

# A versatile 3D stacked vision chip with massively parallel processing enabling low latency image analysis

Stéphane Chevobbe<sup>1</sup>, Maria Lepecq<sup>1</sup>, Karim Benchehida<sup>1</sup>, Mehdi Darouich<sup>1</sup>, Thomas Dombek<sup>1</sup>, Fabrice Guellec<sup>2</sup>, Laurent Millet<sup>2</sup>

<sup>1</sup>CEA LIST Department, CEA LIST, Centre CEA Saclay, Gif-sur-Yvette, France

<sup>2</sup>CEA LETI Department, CEA LETI, MINATEC Campus, Grenoble, France

**Abstract** This paper presents a 2-layer 3D stacked Back Side Illuminated vision chip performing high speed programmable parallel computing by exploiting in-focal-plane pixel readout circuits. The proposed circuit exhibits a 5500fps frame rate, 5 times higher than previous works without reducing ADC resolution. It allows heterogeneous parallel computations on up to 31×31 inter-pixels neighborhoods in a single chip.

## 1. Introduction

Today's embedded systems in cyber physical systems need to process a growing amount of sensor data for high reactivity applications, meaning very low latency between data capture and action. To tackle low latency and high processing demand, designers tend to put the processing units closer to the sensor, creating smart imagers. However, the increase of resolution and framerate in the image sensor implies a bottleneck to output the data to the processing units, degrading the possible reactivity. Smart imagers taking benefit of advanced 3D integration can solve system bottlenecks in various application domains going from very low power IoT solutions to high complexity image based applications under embedded constraints.

Thanks to the 3D stacking technology, new concepts of image sensors are emerging tightly coupling a first layer of backside illuminated (BSI) pixels array and a second layer including digital logic. In the state of the art, this new integration technology is mainly used to increase the quality of the captured image, the acquisition frame rate [1, 2]; or to allow image processing features near the sensor [3], [4]. But the data bottleneck between sensor and processors remains as the image conversion is done through column ADC,

introducing latency while loading processing unit's memories. New architectures investigate parallelism partitioning by assigning one ADC to a group of pixels in order to reach higher frame rate [5] or better data control [6].

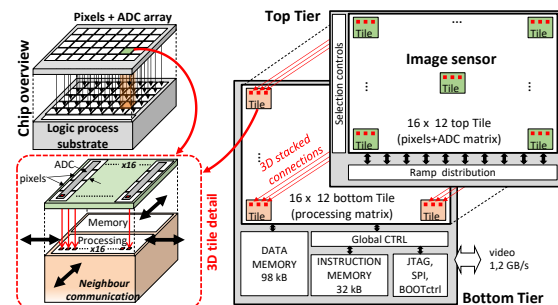


Figure 1 - RETINE architecture and macropixel partitioning

## 2. Smart imager overview

In this work [7], we propose a new approach of 3D stacked imager, named "RETINE": a 2 layers chip based on the replication of a 3D tile in a matrix manner providing a highly parallel programmable architecture as depicted in Figure 1. Each tile is composed of a top tier layer of 16x16 BSI binned pixels array (or 64x64 subpixels array) with associated parallel in-focal-plane pixel read-out circuits and 16 column ADC and a bottom tier layer integrating a homemade SIMD processor able to capture pixels data stream and compute them on the fly at a high framerate.

Thanks to tight coupling and fast feedback, image acquisition parameters such as pixels integration time and digitalization can be locally adapted in each tile.

We focused our design on a general-purpose demonstrator with a high pre-processing flexibility and low latency for high-speed applications. We call demonstrator, the chip implementation of the previously introduced system architecture. Traditional approaches face bottleneck constraint when the sensor is too large. With our approach, this bottleneck issue is removed as data analysis on high dynamic data is performed into the sensor. Moreover, thanks to the in-focal plane vertical connectivity, frame rate is no longer limited by internal data bandwidth and the processing latency is now independent of the image sensor size. Our demonstrator is able to perform data reduction, to ease further high-level processing tasks performed outside the chip. This reduction is enabled by processing tasks like on-the-fly tunable quantization levels, ROI selection, flags dump, or tracking information. The use of 3-D stacked tiled structures allows a scalable implementation for massive parallel computation (Figure 1).

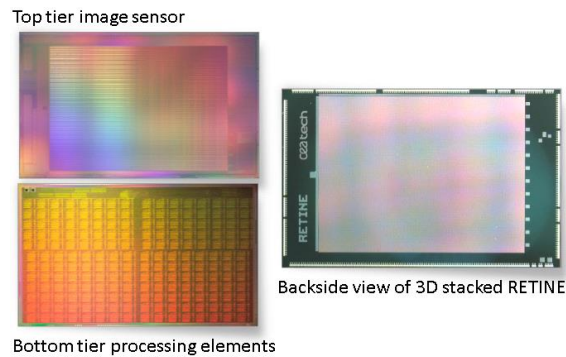


Figure 2 - top and bottom layer of the RETINE chip

The chip supports advanced features such as large distance inter-pixel communication, flexible instruction flow, and a novel differentiated multi-flow execution. In this MPMD mode, different programs are assigned to

several MPXs areas in a parallel execution manner. Up to 4 parallel threads can be natively managed. The microcode memories can also be updated during execution. This mechanism enables the execution of complex applications, composed of several processing kernels alternatively executed thanks to context switch.

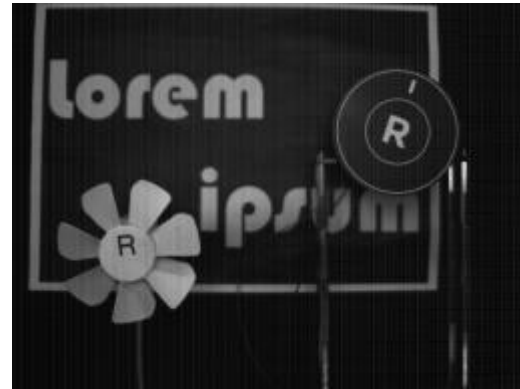


Figure 3 - Snapshot of static scene with RETINE chip @ 1000 fps

RETINE is able to process pixel data during the subsequent acquisition time, and dynamically perform adjustments on the image acquisition flow or decide when to output image features based on the image analysis results.

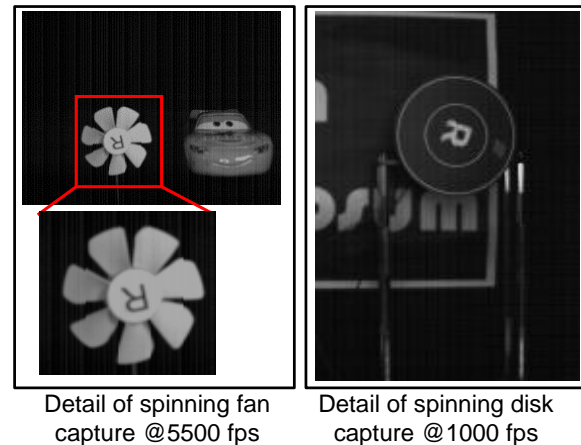


Figure 4 – Snapshot examples of moving scene with RETINE chip

### 3. Circuit characteristics

A prototype of RETINE has been manufactured in CMOS 130 nm with a top pixel array layer of

256x192 binned pixels also addressable as 1024x768 subpixels array coupled with a layer including 16x12 SIMD processors.

Figure 2 shows a picture of the top and bottom layer of the RETINE chip. The chip achieves video rates from 5500 fps in binned mode (cf. Figure 3 and Figure 5) to 340 fps in full resolution mode. The prototype operates at 80 MHz with 720 mW power consumption leading to 85 GOPS/W power efficiency. Figure 4 shows two examples of acquisition speed of fast moving objects such as fans and spinning disc.

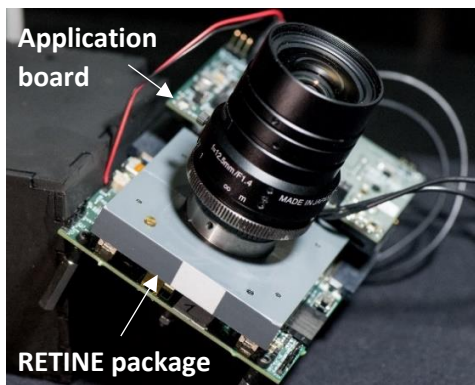


Figure 5 - Application board with RETINE chip

#### 4. Application results

A wide variety of low level image processing kernels have been developed proving the versatility and computing efficiency of the architecture. The RETINE package is plugged on an application board as shown in Figure 5.

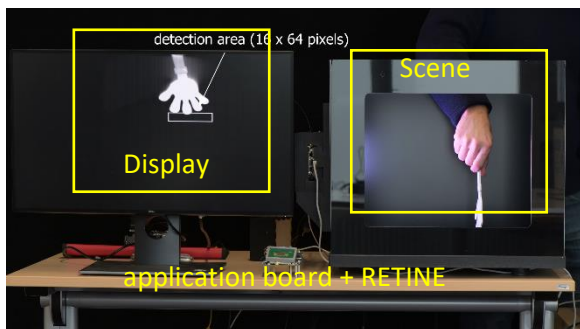


Figure 6 – Demonstration set up with an application board plus RETINE chip, a screen and the box for the moving scene

All the processing is done inside the bottom tier of the RETINE chip. The application board loads the binary codes, feeds the power supplies and displays the results on a screen as shown in Figure 6.

For example, fast event detection (event less than 3 ms) over the whole image has been validated but also event detection restricted to a user-defined area as shown in Figure 6, where the detection area is inside the white rectangle.

Thanks to the high speed acquisition and the tight coupling between the image capture layer and the processing layer, Figure 7 shows 6 pictures of partial images of full frames captured at 1000 FPS. The detail shows a fan rotating at around 8000 rpm. To limit the memory footprint the images are acquired in binary mode in this example.

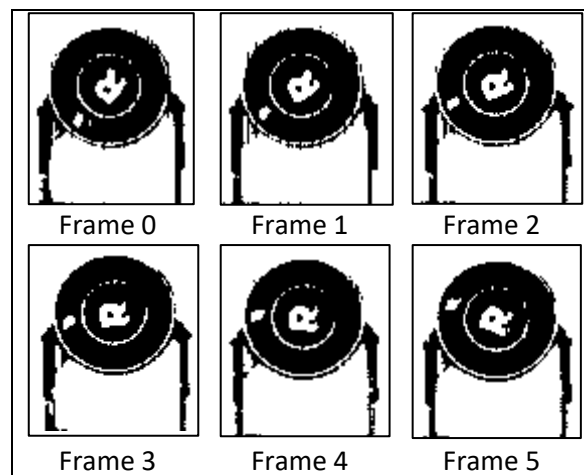


Figure 7 – 6 consecutive binary images from a fan rotating at 8000 rpm captured at 1000 FPS. The 6 pictures are partial images of full captured frames

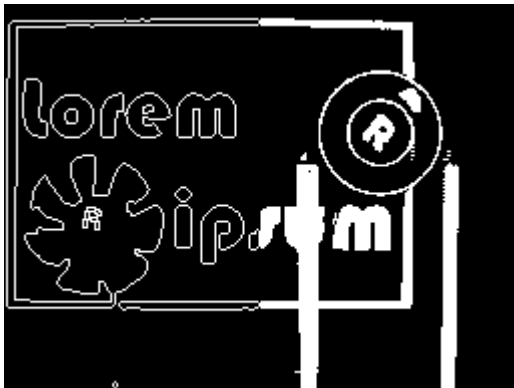
Spatio temporal processing (processing time around 1 ms) such as motion detectors from frame to frame difference have been implemented (cf. Figure 8). In this example the output of the sensor is a frame with three values, 0 for none motion, -1 and +1 for motion. The amount of output data can be reduced by just sending the position and the value of moving pixels. Global processing such as global

histogram and local histogram on multi image areas have also been successfully implemented.



**Figure 8 – Frame difference-based motion detection image from a moving text captured at 1000 FPS**

Finally, the execution of Haar filter and edge detection has proved that RETINE is also well adapted for pixel based feature extraction. On Figure 9, two different processing kernels are executed in parallel : threshold on the half right side and edge detection on the half left side.



**Figure 9 – A 1000 FPS spatial multi-processing capture from a moving scene. This single capture features edge detection on the left side, and threshold on the right side.**

Some others image processing kernels such as labelling, other motion detection or first layers of CNN are also supported and still under work.

## 5. Conclusion

We have demonstrated a 3D stacked vision chip featuring in focal plane readout tightly coupled with flexible computing architecture for configurable high speed image analysis. It paves the way to smart vision chip needed by Cyber Physical System.

## References

- [1] P. Chou, C. Chang, M. M. Mhala, C. C. Liu, C. Y. Chao, C. Huang, H. Tu, T. Wu, S. Yeh, S. Takahashi, and Y. Huang, "A 1.1m-Pitch 13.5Mpixel 3D-stacked CMOS image sensor featuring 230fps full-high-definition and 514fps high-definition videos by reading 2 or 3 rows simultaneously using a column-switching matrix," in 2018 IEEE (ISSCC), pp. 88–90.
- [2] T. Haruta, T. Nakajima, J. Hashizume, T. Umabayashi, H. Takahashi, K. Taniguchi, M. Kuroda, H. Sumihiro, K. Enoki, T. Yamasaki, K. Ikezawa, A. Kitahara, M. Zen, M. Oyama, H. Koga, H. Tsugawa, T. Ogita, T. Nagano, S. Takano, and T. Nomoto, "4.6 A 1/2.3inch 20Mpixel 3-layer stacked CMOS Image Sensor with DRAM," in 2017 IEEE International Solid-State Circuits Conference (ISSCC), pp. 76–77.
- [3] T. Yamazaki et al., "A 1ms high-speed vision chip with 3D-stacked 140GOPS column-parallel PEs for spatio-temporal image processing," in IEEE ISSCC Dig. Tech. Papers., Feb. 2017, pp. 82–83
- [4] O. Kumagai, A. Niwa, K. Hanzawa, H. Kato, S. Futami, T. Ohyama, T. Imoto, M. Nakamizo, H. Murakami, T. Nishino, A. Bostamam, T. Inuma, N. Kuzuya, K. Hatsukawa, F. Brady, W. Bidermann, T. Wakano, T. Nagano, H. Wakabayashi, and Y. Nitta, "A 1/4-inch 3.9Mpixel low-power event-driven back-illuminated stacked CMOS image sensor," in 2018 IEEE International Solid - State Circuits Conference - (ISSCC), pp. 86–88.
- [5] T. Arai, T. Yasue, K. Kitamura, H. Shimamoto, T. Kosugi, S. Jun, S. Aoyama, M. Hsu, Y. Yamashita, H. Sumi, and S. Kawahito, "6.9 A 1.1m 33Mpixel 240fps 3D-stacked CMOS image sensor with 3-stage cyclic-based analog-to-digital converters," in 2016 IEEE International Solid-State Circuits Conference (ISSCC), pp. 126–128.
- [6] T. Takahashi, Y. Kaji, Y. Tsukuda, S. Futami, K. Hanzawa, T. Yamauchi, P. W. Wong, F. Brady, P. Holden, T. Ayers, K. Mizuta, S. Ohki, K. Tatani, T. Nagano, H. Wakabayashi, and Y. Nitta, "A 4.1Mpix 280fps stacked CMOS image sensor with array-parallel ADC architecture for region control," in 2017 Symposium on VLSI Circuits, pp. C244–C245.
- [7] L. Millet, S. Chevobbe, C. Andriamisaina, E. Beigne, F. Guellec, T. Dombek, L. Benaissa, E. Deschaseaux, M. Duranton, K. Benchehida, M. Darouich, M. Lepecq, "A 5500fps 85GOPS/W 3D stacked BSI vision chip based on parallel in-focal-plane acquisition and processing", on Symposium on VLSI Circuits, pp245-256, 2018