A 5 Million Frames Per Second 3D Stacked Image Sensor With In-Pixel Digital Storage

Laurent Millet, Margaux Vigier, Gilles Sicard, Nils Margotat, Fabrice Guelliec, Sebastien Martin
Université Grenoble Alpes, CEA, LETI
F-38000 Grenoble, France
laurent.millet@cea.fr

Wilfried Uhring
ICube Laboratory, University of Strasbourg and CNRS
Strasbourg, France
Wilfried.uhring@unistra.fr

Abstract—A CMOS burst image sensor reaching 5Mfps with 52 frames in-pixel digital memory has been designed and tested. It fully takes advantage of 3D stacked technology to implement a scalable architecture for 8-bits quantization and data storage at pixel level in CMOS technology. This imager also benefits from backside illumination (BSI) for improved fill factor and wide spectrum sensitivity. A demonstrator has been fabricated, embedding two types of 3D based pixel. In this paper we present the very first experimental test results of 3D stacked in-pixel digital burst image sensor. These results show advantages of using 3D technology to obtain a very high frame rate with both relaxed design conditions and readout timing constraint compared to conventional high speed burst image sensors.

Keywords—burst imaging, digital storage, 3D stack, CMOS

I. INTRODUCTION

High speed imaging is a key technology to study high speed phenomenon such as micromechanics, explosives, plasma formation mechanisms or laser ablation. In a classic image sensor (IS), images are acquired with on the fly chip readout. However, for high speed IS this read-out process is a bottleneck limiting the frame rate to about 10 kilo-frame-per-second (fps) for a 1 mega-pixel resolution as presented in [1]. To tackle this limitation, burst IS uses an on-chip memory to store the burst video sequence recorded during the acquisition. Images are then read-out at lower speed. While the frame rate of burst IS can exceed 10 M frames per second (Mfps), they suffer from the small embedded memory limiting the video length to a maximum of hundreds frames.

Burst image sensors are currently divided into two technological categories: CMOS and CCD. CCD imagers can achieve high frame rate with high in-pixel analog memory density [2], [3] and they also benefit from a very high fill factor when using BSI process. However this technology involves several drawbacks. First, they have a huge power consumption to supply the high gate control voltages. Secondly the analog memory is placed in the focal plane. While BSI configuration is much better for fill factor considerations, it exposes the memories to backside photo-current generation which can degrade the stored data. It requires thick substrate to cut wavelengths above 700nm [2], [3]. This in-pixel constraint also limits the memory to 100 frames. In addition the CCD technology impedes the use of embedded functions such as Analog to Digital Converter (ADC), or processing features which arise in smart CMOS solutions e.g. compressive sensing [4].

Up to now, high speed image sensors using CMOS implementation hardly compete with CCD sensors. The CMOS local memorization is also performed thanks to very small capacitors despite large KTC noise. We distinguish two kinds of implementations regarding analog memories: either using memory banks outside the pixel array [5], [6], or using in-pixel memory [7]. In this last case, the main drawback is the very small pixel fill factor. The use of memory banks limits the scalability of the array size and faces power penalty due to signals propagation on the entire array through high capacitance lines at high speed. In both cases, there is also a strong limitation on the readout phase due to leakage on small capacitor memories. They require a fast readout and hence high working frequencies before data deterioration. Some works [4] circumvent this limitation by using time coded multi-capture, but are still limited to 15 frames memory.

A 3D stacked IC approach overcomes these restrictions by allowing a distribution of electronic functions on dedicated tiers [8]. Top tier photodiodes can benefit from imaging technology and backside illumination for a maximal sensitivity, while the bottom tier provide room for in-pixel memories without sacrifying the fill factor (Fig. 1). We designed a scalable 3D pixel using direct bonding stacking, which causes no dead area in the silicon [10] on the contrary of TSV solutions. We also take advantage of this topology to design an in-pixel ADC for embedded digital storage. Our Image Sensor benefits of a 40nm advanced technological node on bottom tier for high density capability. Using RAM-based memory could easily overcome the current hundreds of stored images and allows a video memory of thousand images [9].

This paper describes the architecture and electrical results of the very first 3D-stacked BSI burst image sensor with in-
focal-plane digital memory. The overall chip description is presented in section II. Top tier and bottom tier architecture are described in section III and IV, while experimental results are detailed in section V.

II. OVERALL DESCRIPTION

As said, the 3D stacked implementation is suitable for mixing analog and digital circuits with dedicated technological nodes. Our prototype uses a 90nm image sensor layer over a 40nm digital layer connected by hybrid bonding process. The selected partitioning is illustrated in Fig. 2. The top tier embeds an array of $20\times20$ BSI pixels. Since we need to maximize area for digital memories, each top tier pixel contains a photodiode with its associated readout structure and the analog frontend of a ramp ADC, while the bottom tier pixel contains the counter and digital memory to store 52 images. This one ADC per pixel topology provides a homogeneous layout and avoids fixed pattern noise issues. Note that other partitioning options can be considered in 3 or more layers designs [8]. The bottom chip embeds the control circuits, power distribution, interface controls, and a PLL to drive the counters. The IO ring is also located on bottom tier involving that control signals come from the same die for both top and bottom chips. The top tier is supplied by the bottom chip through 3D Cu-Cu contacts.

III. TOP TIER PIXEL STRUCTURE

A. Photodiode

The 3T BSI photodiode dimensions are $40\mu m \times 50\mu m$ and reach a fill factor of 80%. It is surrounded by a deep trench isolation to avoid inter-pixel interferences and to protect the photosensitive elements from disturbances of the nearby active CMOS circuits (leakage and electron harvesting).

Due to large photodiode dimensions we used multiple collecting contacts. Two different photodiode designs were implemented: A-type version with 4 contacts and B-type version with 15 contacts. These two designs bring different integrating capacitances values (made by junction photodiode and parasitics). A-type shows a lower value (25fF) than the B-type (50fF), but the B-type may exhibit a better time response during charge collection.

B. Pixel readout structure Analog to Digital Converter

The chosen frontend readout structure is a 3T pixel with a sample and hold stage, selected for its fast time response. A CTIA could be implemented for better voltage control on sensing node and signal gain thanks to feedback capacitances. Yet, since the photodiode capacitance is expected to be low, the source follower has been considered as a better option.

A sample and hold circuit is used to stop integration and ensure a clean signal for the ADC. The acquired signal is then quantized while next frame is already integrating on photodiode capacitance.

C. Analog to Digital Converter

The ADC topology is a ramp converter. The comparator is designed in the top tier (Fig. 3) while the counter circuit is in the bottom layer. The comparator is a double stage amplifier followed by a level shifter for voltage compatibility with digital layer. It performs a correlated double sampling on pixel signal, and embeds a double offset cancellation circuits through AZ1 and AZ2 stages. Note that a comparator usually produces kickback noise on reference voltage lines and power supply, magnified by the matrix dimensions. This is an issue at our operating frequencies, since the references take time to come back to their initial level. To tackle this problem we implemented a clamp circuit to drift current even after the comparator triggering, which ensure a constant power consumption over the matrix array and brings a significant kickback noise reduction.

Due to 3D connections per pixel. The first is the comparator output signal called “stop_conv”, from the top to bottom. It freezes the counter value. The second one comes from the bottom chip. It brings the 1.2V power supply to the top tier for the level shifter.

D. Top tier pixel overview

The layout implementation of an A-type pixel and the readout circuits is shown in Fig. 4. A metal 1 and 2 shield is drawn over the photodiode, and the BSI implementation allows to use free space for fringe capacitances. One of the 3D connections hangs over the neighbor pixel.
IV. BOTTOM TIER PIXEL STRUCTURE

The bottom tier pixel embeds counters for the ramp ADC, a bank of 52 Bytes memory, and top/bottom synchronization circuits.

A. Analog to Digital Conversion

Digital pixel comprises an 8 bit counter, driven by a clock up to 2.56GHz generated by a peripheral PLL circuit. This frequency ensures an 8bits conversion at 5Mfps. When the conversion begins, the counter value is incremented at each clock rise and stops as soon as the top tier comparator has switched. The digital value is then automatically stored in the selected 13 bytes FIFO out of the 4 available per pixel (Fig. 5). This implementation enables continuous recording while no event occurs. This is convenient if the imager is waiting for a specific event, since the acquisition can be stopped with a dedicated external trigger. Data are then read-out through a low frequency serial link. Each pixel memory can be selected through “Row_SEL” gate. There is no time constraint here since digital memories do not suffer from leakage issues like in CCD or CMOS analog storage solutions.

B. Signal control and synchronization

Bottom tier handles synchronization signals. Since the frame capture is global shutter based, the 20×20 pixels of the array work in parallel. Hence, 3D drive signals from bottom to top tier are sent outside the focal plane for line-broadcast communication, as well as ramp distribution. Start conversion signals are sent to top and bottom tier outside pixel array for conversion synchronization of both layers as depicted in Fig. 6.

C. Embedded Digital Memory

For this first prototype digital memories are realized with D Flip-Flops for a total memory area of 45μm × 45μm. This memory topology allows 52 frames of 8 bits to be sequentially written. A much larger memory depth could be achieved by using a high density SRAM instead of D-Flip Fops. It would involve an important increase of storable frames. Typically it could allow around 775 frames in 28nm technology and 387 frames in 40nm technology for the same pixel pitch while including the SRAM control circuit. Implementing SRAM will involve partitioning issues (one cut for several pixels) and local pixel addressing considerations. For instance this SRAM partitioning has already been investigated in the context of high speed imaging [8].

V. ELECTRICAL MEASUREMENTS

To evaluate our demonstrator performances, two kind of measure analysis have been performed. First photodiodes have been characterized thanks to a test structure allowing access to the internal nodes. Secondly we also demonstrated standard full frame electrical acquisition.

A. Photodiode characteristics

Photodiode characterization measures were performed on an optical bench including a monochromator. The extracted spectral sensitivity is shown in Fig. 7. Since we are not limited by CCD constraints, i.e. the use of thick oxide to prevent large wavelength from corrupting the analog memories, our measured diagram exhibits a wide spectrum range from 360nm to 900nm. The graph oscillations appearing between 600nm and 900nm are due to constructive and destructive reflection interferences in silicon interfaces and known as the etalon effect.

B. Dynamic captures: pulse response

We performed several transient measures to evaluate the IS response to fast moving scenes. On the first example shown in
**TABLE I. TECHNICAL SPECIFICATIONS SUMMARY**

| Scalable Pixel specifications | Readout characteristics |
|------------------------------|-------------------------|
| Technology                   | Read Noise | Full well capacity | Dynamic range | 1 V |
| 4M1P 90nm / 7M1P 40nm        | 0.59 mV rms | 137 ke-            | 7.3 µV/e-     |
| Pixel size                   | 50µm×50µm            | 20                  |
| Fill factor                  | 80%                   | 50                  |
| Frame rate                   | 5 Mfps                | 7.3                |
| Memory                       | 52 (8 bits)           | Chip Considerations |
| depth (frames)               | (387 in 40nm SRAM)*   | Pixel count        | 20 x 20 |
|                             | (775 in 28nm SRAM)*   | chip size          | 2.7 x 2.7 mm |

* estimation

Fig. 8. Two 3 frames sequences to evaluate beam persistence on A-type and B-type pixels (left), and the test bench (right)

On the first acquisition (stream A) a single pulse focused on the upper side of the imager is sent during the frame 2 to test the A-type pixel response. The laser spot is surrounded by a diffraction circle. The high illuminated central spot causes a small residual remanence illustrated by a 20 codes dot visible on the frame 3 (red arrow). This effect is not visible when the B-type pixel is illuminated on the lower side of the imager (stream B). It can be explained by the fact that the A-type pixel has a small amount of residual drifting charges due to insufficient collecting contacts, which is not the case for the B-type photodiode. The circuit exhibits a pixel aperture ratio of 73.7%, measured thanks to the 100ps FWHM light pulse.

**C. Dynamic captures: sine response**

This second example evaluates the response to a half-sine wave exposition. 5 different acquisition streams are made at 5Mfps. Each capture is slightly shifted by 200ns from the previous one, and is presented on Fig. 9. For a better estimation of the captured signals, a graph shows the median value of each photodiode type (A or B) for each image. We clearly see that the B-type photodiode is less sensitive than the A-type variation. This is due to the higher junction capacitance, hence a higher full well capacity. However in both cases each sine capture is clearly shifted and we demonstrated that the process is reliable and replicable. Several pictures of the first capture stream are displayed for assessment.

**VI. CONCLUSION**

In this paper we presented electrical results of the very first burst image sensor taking advantage of 3D stacked technol-