A RGBZ $1.4\,\mu\mathrm{m}$ Image Sensor for Color and Near-Infrared Indirect Time-Of-Flight Depth Sensing on Monolithic Silicon

Clémence Jamin-Mornet[1], Hugo Dewitte[1], William Guicquero[1], Gaëlle Palmigiani[1], Olivier Saxod[1], Lionel Gerard[1], Cyril Bellegarde[1], Laurent Coindoz-Bernard[1], Pascal Fonteneau[2], and Arnaud Tournier[2]
[1]*Université Grenoble Alpes, CEA-Leti, F-38000 Grenoble, France, email: clemence.jamin@cea.fr*
[2]*STMicroelectronics, 850 rue Jean Monnet, Crolles, 38926, France*

***Abstract*—A 1.2 Mpix RGB+Z image sensor with $1.4\,\mu\mathrm{m}$ pixels has been designed to simultaneously capture RGB color images with accurate color reproduction and parallax-free depth maps on a single chip. This paper presents the design and structure of this monolithic silicon-based co-integration of $1.4\,\mu\mathrm{m}$ RGB pixels with 2-tap iToF pixel (2x2 binning), fully isolated by deep trench isolation, and an on-chip optical filtering. The Quantum Efficiency (QE) of RGB pixel is close to the state-of-the-art and with a high rejection at 940 nm. The Z-pixels, taking advantage of a $5.8\,\mu\mathrm{m}$ gradually doped epitaxial layer, achieve state-of-the-art QE at 940 nm (when compared to other RGBZ systems) and a 78 % demodulation contrast at 200 MHz.**

***Index Terms*—RGBZ, Time-Of-Flight, Demodulation Contrast, color reproduction, monolithic sensor, parallax-free, BSI**

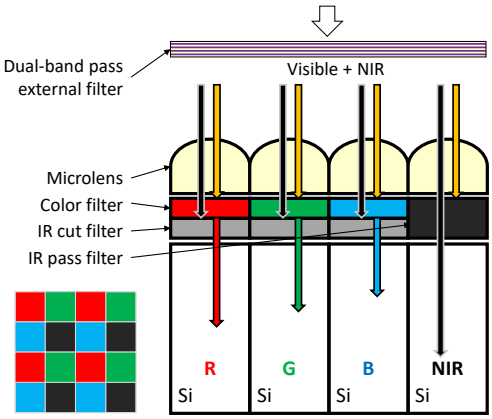


Fig. 1: Structural overview of the proposed RGBZ monolithic integration

I. Introduction

In recent years, the need for advanced imaging systems combining both high resolution color image (2D) and ranging information (3D) has emerged, driven by applications such as robotics, automotive, security, virtual reality, or improved bokeh effect for smartphones. Current commercial 2D/3D imaging solutions use two separate sensors to capture the color image and depth, leading to misalignment between the color frame and the depth map. In [1] and [2], a time-division multiplexing architecture is proposed; however, this solution limits the frame rate. In 2012, Kim et al. introduced the first RGBZ sensor that simultaneously captures both color and depth information on the same chip with two RGB and Z interlaced arrays [3]. However, some limitations can be highlighted, such as NIR pollution into visible pixels and non-uniform RGBZ array, which lead to poor color quality. Stacking solutions have also been proposed, leading to compromises on RGB image quality. Indeed, stacking silicon RGB pixels on silicon NIR pixels results in NIR color mixing and poor color reproduction [4] while stacking organic photoconductive RGB film on silicon iToF pixel faces the current limitations of organic performances on visible pixels compared to state-of-the-art silicon-based pixel performances [5]. This paper presents a novel image sensor that combines color (RGB) and depth (Z) sensing in a monolithic design, offering a compact, parallax-free, low-power, and system-level cost-effective solution for future imaging technologies. It overcomes the traditional limitations of the silicon monolithic architecture by implementing on-chip optical filtering of the NIR wavelength above color pixels and a uniform RGBZ pattern at $1.4\mu m$ pixel pitch with 2x2 binning of Z pixels (Fig. 1). Moreover, combining the reduction of the Z pixel pitch with advanced image reconstruction algorithms allows to limit the image quality degradation due to the introduction of depth pixels in the RGB pattern.

II. Device Architecture

A. Pixels Architecture

Fig. 2 illustrates the simplified layout and the pixel circuit of the RGBZ imager. The image sensor is developed in a 140 nm technology node on a 2D test chip with fully isolated Back-Side Illuminated (BSI) pixels by capacitive deep trench isolation (CDTI) technology. The architecture can be easily adapted to a 3D-stacked configuration.

The $1.4\mu m$ RGB pixels are based on a 4T-like structure with shared transistors. They are grouped in sets of six pixels; six photodiodes with their own Transfer Gate (TG) are connected to a common readout circuit composed of the

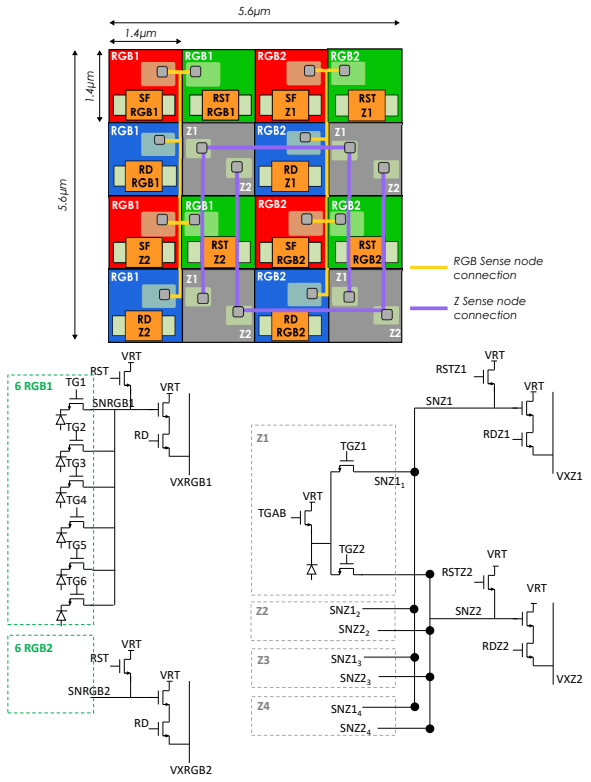


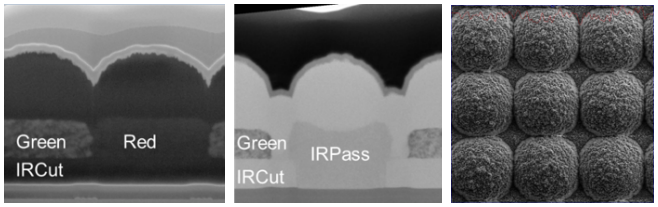
Fig. 2: Schematics of RGBZ layout and pixel circuits of RGB and 2-tap iToF pixels.

standard Reset (RST), Read (RD), and Source Follower (SF) MOSFETs, distributed over the visible pixels.

The $1.4\,\mu m$ Z pixel are 2-tap iToF pixels with static drift field pinned-photodiode and low-noise buried-channel transfer gates as reported in [6]. The pixels are composed of an anti-blooming gate (TGAB), two high-speed transfer gates (TGZ1 & TGZ2) with two memories, Z1 and Z2. Each sense node SNZ1 and SNZ2 is connected by macropixels of 2x2 Z pixels (four SNZ1 and four SNZ2 connected together). These four Z pixels share a common reading circuit for each Z1 and Z2 path, composed of the Read (RDZ1 & RDZ2), Reset (RSTZ1 & RSTZ2) and Source Follower (SFZ1 & SFZ2) conventional MOSFETs. Because of this architecture, the Z pixels operate in voltage domain-like mode, which implies kTC noise due to the 3T-like readout scheme. The Z pixels are "demodulation" pixels, also called lock-in pixels. They are used for a continuous-wave indirect time of flight (CW-iToF) system featuring a high-frequency laser modulation (typically from 10 to 300 MHz) for the active illumination of the scene at a wavelength of 940 nm. The measured distance can be expressed as in Eq. 1 for a 2-tap iToF pixel with 2 captures (4-bin demodulation).

$$depth = \frac{c_{light}}{4\pi f_{mod}}\varphi\ ,\ \ \varphi = tan^{-1}\left(\frac{A_1 - A_3}{A_2 - A_0}\right) \qquad (1)$$

where f_{mod} is the modulation frequency of the source; A_0, A_1, A_2 and A_3 are the signal accumulated in the two memories, A_1 and A_2 from Z1 and A_0 and A_3 from Z2. Therefore, for a typical 10 to 300 MHz operating range, the



(a) (b) (c)

Fig. 3: (a) FIB SEM horizontal cross section of color pixels (b) TEM diagonal cross section of Green-Depth pixels (c) Top view of $1.4 \times 1.4\,\mu m$ microlens before top coat

Z pixel can measure a theoretical maximum distance range of 15 to 0.5 meters.

B. Device Technology

For standard color image sensors, an external IR filter with a cut-off at 650 nm is typically used. In the case of a RGBZ image sensor, where the depth pixel operates with continuous-wave active illumination at 940 nm, the intrinsic Near-InfraRed (NIR) light pollution for both RGB and Z pixels needs to be filtered using an external dual-bandpass filter. This filter consists of a broad band in the visible wavelength range and a narrow band at 940 nm. Additionally, an IR-pass black resin is patterned on the Z pixel to block the visible wavelengths, minimizing visible light pollution during depth measurements at 940 nm. On the RGB pixels, an IRCut resin is patterned below the standard color resins to suppress remaining NIR pollution in the visible spectrum in order to preserve the color image quality. Fig. 3 shows scanning and transmission electron microscope (SEM / TEM) images of the optical stack on the fabricated device with on-chip pixelated filtering solution and micro-lens.

The thickness of the IRCut resin is a critical parameter for the RGBZ performance, and it has been adjusted through optical simulations. The objective is to filter NIR light while limiting the absorption of the visible wavelength. In addition to simulation, various resin thicknesses have been tested through splits on a processed CMOS lot.

III. Characterization Results

A. RGB pixels

The RGB pixels performance is reported in Tab. I and compared to other works. The RGB pixels have a Q_{sat} of 5800 electrons with a conversion gain of $65\,\mu$V/e-. The Quantum Efficiency (QE) of RGB pixels, illustrated in Fig. 4, shows low residual content at 940 nm while maintaining high efficiency in the visible spectrum, thanks to the IRCut resin. The ideal performance of this resin has been characterized on a standard Bayer pattern with unpatterned IRCut resin, resulting in a rejection below 1% at 940 nm, which is assumed to be the performance of the IRCut in an RGBZ pattern without optical crosstalk from Z to RGB pixels. In this work, optical crosstalk is the most limiting factor on the QE performance of RGB pixels. Far-Back-End Optic (FBEO) improvements, such as isolation between

	Parameters	Unit	This work	IEDM 2024 [5]	ISSCC 2012 [3]
	Sensor material	/	RGB: silicon Depth: silicon (940 nm)	RGB: organic Depth: silicon (NIR)	RGB: organic Depth: silicon (850 nm)
RGB Pixels	Pixel pitch	μm x μm	1.4 x 1.4	1.0 x 1.0	2.25 x 2.25
	Full well	e-	5.8k	9.8k	/
	QE Green@520 nm	%	normalized	59%	/
	QE Green@940 nm	% / a.u.	$<$ 5 a.u. (norm 100 @520nm)	0%	/
Z Pixels	Pixel pitch	μm x μm	1.4 x 1.4 (2.8 x 2.8 binning)	4.0 x 4.0	2.25 x 9.0
	Number of taps	tap	2	2	2
	Temporal Noise @ 40°C	e-	38	/	/
	QE @940 nm	%	normalized	19%	/
	Pulsed PLS@10 MHz	%	3%	/	/
	DMC @200 MHz @40°C	%	78%	80%	/

TABLE I: Key performances summary table.

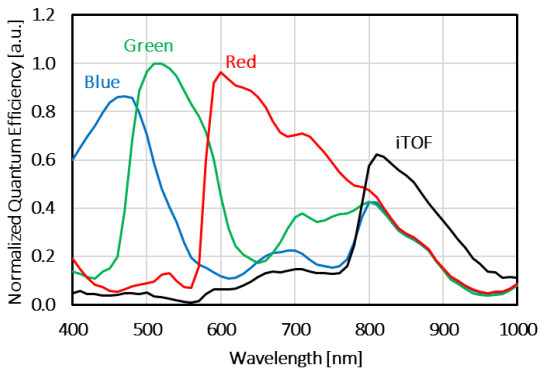


Fig. 4: Normalized QE spectra of RGB and iToF pixels.

color, IRCut, and IRPass filters by tungsten walls or air gaps [7], could therefore be considered to achieve high-quality RGB images.

B. Depth pixels

The Z-pixel performance is also reported in Tab. I. For iToF pixels, the main goal is to maximize the Quantum Efficiency at 940 nm while maximizing the demodulation performance, – *i.e.*, the ability of the Z pixels to transfer the photogenerated charges to the correct memory (Z1 or Z2) – thus maximizing their accuracy in measuring the distance. Two metrics are used to characterize the demodulation performance of the depth pixels, the Impulse Response Time (IRT) and the Demodulation contrast (DMC). The IRT measures the demodulation performance of the pixel for a very short laser pulse ($\sim$ ps), while the DMC is the same but for a sine wave, as it would be the case during nominal system operation. The IRT and DMC measurements are presented in Fig. 5 for frequencies ranging from 10 to 300 MHz. A thick epitaxial layer is required to increase the QE in the NIR wavelengths but degrades the IRT/DMC at high frequencies as the photo-generated electrons take longer to reach the top of the photodiode. The proposed pixel takes advantage of a $5.8\,\mu m$ thick epitaxy layer but gradually doped, with a deep diode to maximize the vertical drift field in order to improve the high frequency response. With this, a state-of-the-art QE at 940 nm (compared to other RGBZ solutions) is reached while preserving acceptable performance at high frequencies with a calculated DMC of 78% at 200 MHz. To further improve the QE at 940 nm, a wavelength diffraction structure could be used at the silicon interface [8] as well as an improved broadband antireflective layer. Furthermore, fine-tuning the doping profiles would be required to reach state-of-the-art iToF performance.

IV. Image Acquisition and Rendering

Fig. 6 reports a 2D/3D image rendering. The acquisition was performed under a laboratory incandescent ambient light source (with high NIR light content) for the RGB image and with an active 140 MHz-modulated light source at 940 nm for the depth image. The reconstruction, which does not require any parallax correction, is performed by a full-custom image rendering pipeline. Our RGBZ rendering software is based on a deep learning CNN model, including depth super-resolution, color demosaicing, adjustable white balance, and gamma correction. The model includes pixel-wise gating with U-Net-generated control signals, promoting a local attention-like mechanism while maintaining a reasonable size ($<$100k parameters). It was trained using artificial data inputs, *i.e.*, simulated mosaiced images and downsampled depth maps, with additional data augmentation, including noise and localized bad depth estimations. Moreover, residual row-to-row RGB input raw data artifacts have been locally mitigated to generate Fig 6b. Finally, for Fig 6c, the initial low-resolution depth map is generated through a phase estimation with 4-bin iToF reconstruction averaged three times (with three phase shift).

V. Conclusion and Perspectives

In conclusion, this paper demonstrates the feasibility of a monolithic silicon-based co-integrated RGBZ image sensor with solid performance for both RGB and Z pixels. In addition, it has been shown that the loss in color image quality due to the occlusion introduced by the Z pixels in the RGB pattern can be compensated for by advanced image reconstruction algorithm. However, efforts are still needed to match the performance of state-of-the-art stand-alone RGGB and iToF image sensors. Several technological improvements could be implemented, such as; the isolation walls between color and IR filters to improve optical crosstalk, fine-tuning

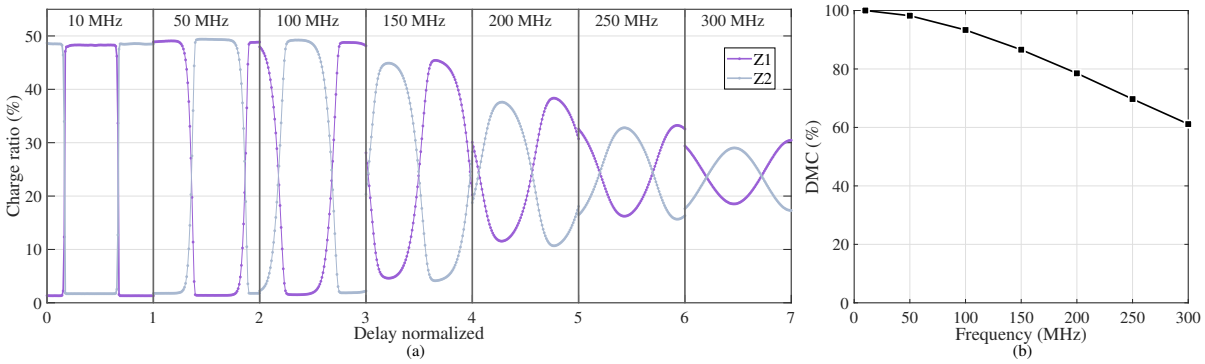


Fig. 5: (a) Charge ratio in Z1 and Z2 memories for different delays of the laser pulse. (b) Evolution of the DMC calculated form the IRT for frequencies up to 300 MHz.

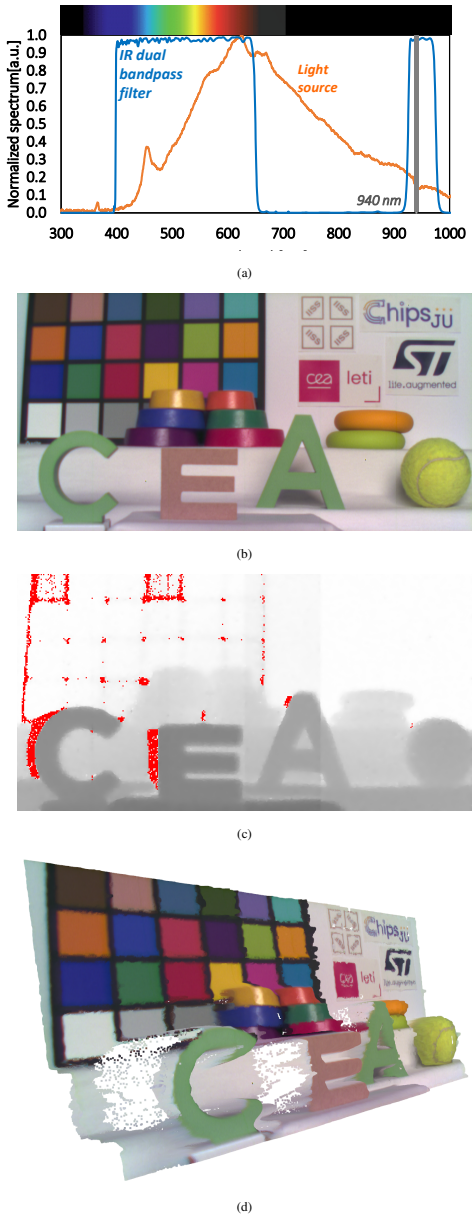


Fig. 6: Image rendering from an acquisition by the $1.4\mu m$ RGB+Z image sensor (a) Illumination conditions with light spots and external dual bandpass filter (b) RGB image reconstruction (c) Depth image with 2-tap $1.4\mu m$ Z pixel at 140 MHz (red pixel are low amplitude measurements to discard) (d) 3D+RGB point cloud reconstruction

the epitaxy doping profiles to improve the DMC of Z pixels, or an improved broadband antireflective layer and a diffraction structure at the silicon interface to increase the Z pixel quantum efficiency. With these improvements, monolithic RGBZ image sensors could approach state-of-the-art performance for both RGB and Z pixels, while still benefiting from the monolithic construction – *i.e.*, a compact, parallax-free, cost-effective and low-power all-in-one sensor.

VI. Acknowledgment

This project has received funding from the CHIP JU program under grant agreement N°101139941. This communication reflects the views only of the author, and the Commission cannot be held responsible for any use, which may be made of the information contained therein. The authors gratefully acknowledge the contributions of the design and process integration teams at STMicroelectronics and the platform at CEA-Leti for their contribution and support.

References

[1] S. Machida *et al.*, "A 2.1Mpixel organic-film stacked RGB-IR image sensor with electrically controllable IR sensitivity," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 78–79.

[2] S.-J. Kim *et al.*, "A CMOS image sensor based on unified pixel architecture with time-division multiplexing scheme for color and depth image acquisition," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 11, pp. 2834–2845, 2012.

[3] W. Kim *et al.*, "A 1.5Mpixel RGBZ CMOS image sensor for simultaneous color and range image capture," in *2012 IEEE International Solid-State Circuits Conference (ISSCC)*, 2012, pp. 392–394.

[4] Y. Takemoto *et al.*, "Multi-storied photodiode CMOS image sensor for multiband imaging with 3D technology," in *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, pp. 30.1.1–30.1.4.

[5] T. Ohkubo *et al.*, "A color image sensor using $1.0-\mu$m organic photoconductive film pixels stacked on $4.0-\mu$m si pixels for near-infrared time-of-flight depth sensing," in *2024 IEEE International Electron Devices Meeting (IEDM)*, 2024, pp. 1–4.

[6] C. Tubert *et al.*, "4.6µm low power indirect time-of-flight pixel achieving 88.5% demodulation contrast at 200MHz for 0.54Mpix depth camera," in *ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC)*, 2021, pp. 135–138.

[7] H. Bak *et al.*, "Advanced color filter isolation technology for sub-micron pixel of cmos image sensor," in *2022 International Electron Devices Meeting (IEDM)*, 2022, pp. 37.6.1–37.6.4.

[8] S. Yokogawa *et al.*, "IR sensitivity enhancement of CMOS Image Sensor with diffractive light trapping pixels," *Sci Rep*, vol. 7, no. 1, p. 3832, Jun. 2017, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-017-04200-y