An 8.7Mpixel 240fps CMOS Image Sensor with 4 × 4 Pixel-Block Local Adaptation of Resolution, Frame Rate, and Exposure Time for Scene Adaptive Imaging

Kohei Tomioka[1], Kodai Kikuchi[1], Akira Honji[1], Takenobu Usui[1], Kazuya Kitamura[1], Shoji Kawahito[2]
1 NHK Science & Technology Research Laboratories, Tokyo, Japan
2 Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan
TEL: +81-3-5494-3326 E-mail: tomioka.k-dk@nhk.or.jp

***Abstract*— An 8.7Mpixel CMOS image sensor with local adaptation of imaging parameters for each 4 × 4 pixel block is presented. The 3,904 × 2,224 pixel array is divided into 976 × 556 blocks. The resolution, frame rate, and exposure time are adjustable per block using shared floating diffusions and pass transistors of charge transfer pulse for 2 × 2 photodiodes, enhancing image quality while suppressing data rates. The device achieved 4K resolution, 240 fps, and 88 dB dynamic range.**

I. INTRODUCTION

Wide-field-of-view videos for immersive media, such as 360-degree video, have attracted attention for commercial use and scientific analysis [1]. A single frame in these videos can contain a variety of detailed patterns, various movements, and varying brightness levels compared to those in conventional narrow field-of-view videos, requiring image sensors with higher resolution, faster frame rate, and high dynamic range. However, these characteristics are subject to tradeoffs due to the chip design constraints, including pixel signal readout speed, power consumption, and chip size [2]. Moreover, increasing the resolution and frame rate leads to higher output data rates, resulting in larger systems and increased power consumption [3],[4]. We propose a CMOS image sensor (CIS) that can locally and adaptively adjust the resolution, frame rate, and exposure time for each 4 × 4 pixel block for scene adaptive imaging. Allocating the necessary imaging performance to the required local area based on subject characteristics improves the image quality while suppressing data rates. Several image sensors have been proposed for the local adaptation of imaging parameters, including exposure-time adjustment for a 32 × 32 pixel block [5], exposure time and frame rate adjustment for a 16 × 16 pixel block [6], and exposure time and frame rate adjustment for individual pixels [7]. These sensors require a relatively large number of pixels per pixel block or a large pixel architecture owing to the local control circuitry. CISs with resolution control have been reported, but local adaptation is not supported [8], or the number of pixels per pixel block is as large as 64 x 64 pixels [9]. Our CIS achieves local adaptation of imaging parameters through two key architectures: (i) A 4 × 4 pixel block with shared floating diffusions (FD) for 2 × 2 photodiodes (PDs) and pass transistors for each PD, which control charge transfer to the shared FDs of the 2 × 2 PDs, and (ii) a control circuit that generates different pass transistor control signals for each pixel-block column in synchronization with the readout scan.

II. SENSOR ARCHITECTURE

Figure 1 illustrates the block diagram and the pixel architecture of the CIS. The 3,904 (H) × 2,224 (V) pixel array featuring 2.5 μm pixels is divided into 976 (H) × 556 (V) blocks, each consisting of 4 × 4 pixels. Each set of 2 × 2 adjacent pixels, i.e., PD(A), PD(B), PD(C), and PD(D), shares an FD and a source follower (SF). The pixel driver scans one pixel block row at a time by generating RT, TX, and SL pulses in the selected row. The four SF outputs from each pixel block are connected to four vertical signal readout lines (PO1, PO2, PO3, and PO4) and converted to digital signals in parallel using four column-parallel readout circuits composed of programmable gain amplifiers (PGAs) and analog-to-digital converters (ADCs). The transfer gate node is shared by each of the four PDs—PD(A), PD(B), PD(C), and PD(D)—within the pixel block. Pass and pull-down transistors are employed for each transfer gate node. They enable or disable TX pulses for each PD based on the TX control signals (TXC(A), TXC(B), TXC(C), and TXC(D)) generated by the mode controller. When TXC(A) is on and TXC(B), TXC(C), and TXC(D) are off, the TX pulse is enabled for PD(A) and disabled for the other PDs. Conversely, when all the TXC signals are on, the charges from PD(A), PD(B), PD(C), and PD(D) are summed and transferred to the shared FD, facilitating 2 × 2 pixel binning.

Figure 2 illustrates the readout methods for the four imaging modes and assignment methods for each pixel block. The readout scans of the pixel driver are performed in cycles of 1/240 s. The readout period for a pixel block row is 1H = 7.26 μs. The TX is turned on twice for each pixel-block row: once for the readout pulse (operates when the SL is on) and once for the electric shutter pulse (operates when the SL is off). The TXC(A), TXC(B), TXC(C), and TXC(D) signals, for which different signals are generated for each block column by the mode controller and wired vertically for each block column, enable the following operations for each imaging mode. In the Normal mode, a subpixel readout technique is utilized where the readout pulses for pixels A, B, C, and D are sequentially enabled in each readout scan of 1/240 s. This results in a readout each pixel with an exposure time of 1/60 s and a frame rate of 60 frames per second (fps). In the Bright mode, a subpixel readout is performed, and an electronic shutter pulse is enabled for all the pixels in each readout scan. This limits the exposure time to 1/240 s. In the Low-Light mode, subpixel readout and four readout skips are performed alternately. This reduces the frame rate to 30 fps and increases the exposure time to 1/30 s. In the Fast mode, readout pulses for pixels A, B, C, and D are enabled in each

readout scan, resulting in a pixel-binning readout of 2 × 2 pixels with an exposure time of 1/240 s and a frame rate of 240 fps. The control pattern of the TX control signals generates different signals for each pixel-block column and is updated by the feedback signal input (38-ch 13.5-MHz SPI signal) synchronized with the pixel-block row selection by the pixel driver. Similarly, the mode controller can set the PGA gain to either ×0.5 or ×2 for each block column. This enables the assignment of different imaging modes and PGA gains for each pixel block using the feedback signal.

III. Experimental Result

Figure 3 shows the specification and the imaging characteristics of each imaging mode. The sensitivity differs among the modes owing to differences in the exposure time, enabling/disabling of pixel binning, and PGA gain. These parameters can be aligned by multiplying the sensitivity correction factor K, which is to be performed off-chip. In Bright mode, the exposure time is reduced to a quarter, so K is 4, and in Low-Light mode, the exposure time is doubled, so K is 0.5. The effective imaging characteristics can be utilized by changing K for the low random noise and high full-well capacity (FWC). Thus, the effective random noise is 1.2 e- in the Low-Light mode, and the effective FWC is 31,000 e- in the Bright mode. The dynamic range can be increased by switching among the Low-Light, Normal, and Bright modes based on the brightness of the subject in each pixel-block. For moving subjects, high-speed imaging at 240 fps can be achieved by using Fast mode. In this mode, the output data rate is suppressed through 2 × 2 pixel binning. In Fast mode, dynamic-range expansion can be realized by switching the PGA gains. Since the pixel binning causes a signal four times the pixel saturation level in maximum to be output from the SF, a gain of 0.5 is applied to the high-luminance signals in the PGA to keep them within the ADC conversion range.

Figure 4 shows the acquired images with local adaptation for each 4 × 4 pixel block. The diagram on the bottom left shows the mode set for each 4 × 4 pixel block. High-brightness areas are set to the Bright mode (green), rotating object areas with a shape of circle are set to the Fast mode (red), dark areas inside the boxes are set to the Low-light mode (blue), and the other areas (black) are set to Normal mode. Overexposure (a), underexposure (b), and motion blur (c) are improved by local adaptation of the imaging parameters. It also shows that the local adaptation of the 4 × 4 pixel block allows appropriate mode control tailored to the contours and details of the subject.

Figure 5 depicts the chip characteristics. The CIS has two operation modes. (i) local adaptation of 4 × 4 pixel block: This mode supports frame rates of 240/60/30 fps, features a multi-sampling 14-bit ADC, and includes noise suppression through the multisampling ADC and digital correlated double sampling (CDS) [10]. (ii) local adaptation of 8 × 8 pixel block local adaptation: This mode supports frame rates of 480/120/60 fps and uses a 12-bit ADC without noise suppression. In this mode, analog CDS is performed in the PGA circuit. The CIS has a die size of 14.2 × 17.8 mm^2 (pixel area: 9,760 × 5,560 μm^2) and consumes 2,640 mW of power.

Figures 6 and 7 compare the performance with previously reported CISs [3-8, 11, 12], and Figure 8 shows a chip micrograph. Compared to other sensors, our CIS offers highly localized imaging parameter control at a high spatial resolution of 4 × 4 pixel blocks, with the smallest reported pixel block size. In addition to local adaptation of exposure time and frame rate, our CIS also supports resolution control, which contributes to suppressing the output data rate. As a result, it achieves a well-balanced performance with 8.7 Mpixel resolution, 240 fps, and a dynamic range of 88 dB, outperforming others in both FoM1 and FoM2 metrics. The proposed CIS architecture successfully addresses the trade-offs among resolution, frame rate, and dynamic range, making it highly effective for enhancing the image quality of wide field-of-view videos, which is an essential requirement for immersive media.

IV. Conclusion

This paper presented a CMOS image sensor capable of locally adaptive control of resolution, frame rate, and exposure time for each 4 × 4 pixel block. The proposed architecture achieves 4K resolution, 240 fps, and an 88 dB dynamic range while suppressing output data rates, enabling high-quality and efficient imaging, which is crucial for immersive media.

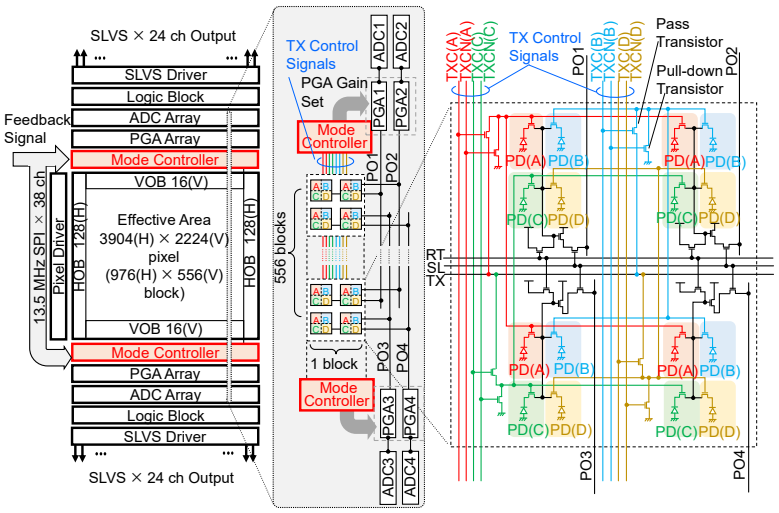


Figure 1: Block diagram and pixel architecture.

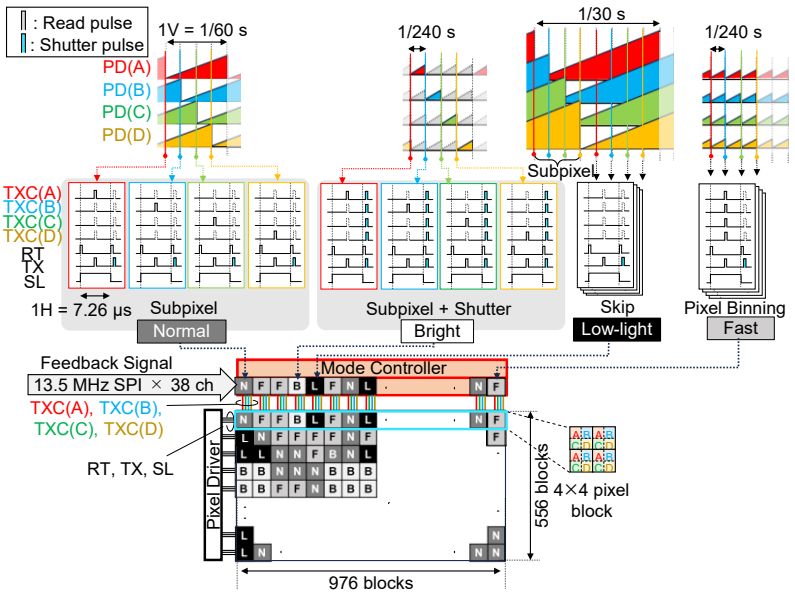
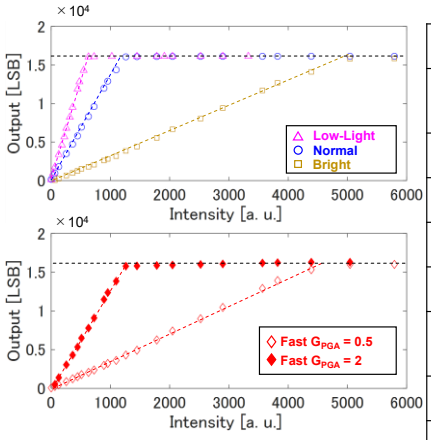


Figure 2: Readout methods for the four imaging modes and assignment methods for each pixel block.



Mode	Normal	Bright	Low-Light	Fast	
Pixel Number Per Block (pixels)	4 × 4	4 × 4	4 × 4	2 × 2	
Sampling Pitch (µm)	2.5	2.5	2.5	5	
Frame Rate (fps)	60	60	30	**240**	
Exposure Time (s)	1/60	1/240	1/30	1/240	
PGA Gain G_{PGA}	2	2	2	0.5	2
Sensitivity Correction Factor K	1	4	0.5	4	1
Random Noise (e^-) × K	2.4	9.8	**1.2**	9.8	2.4
FWC(e^-) × K	7,750	**31,000**	3,880	**31,000**	7,750

Figure 3: Characteristics of each imaging mode.

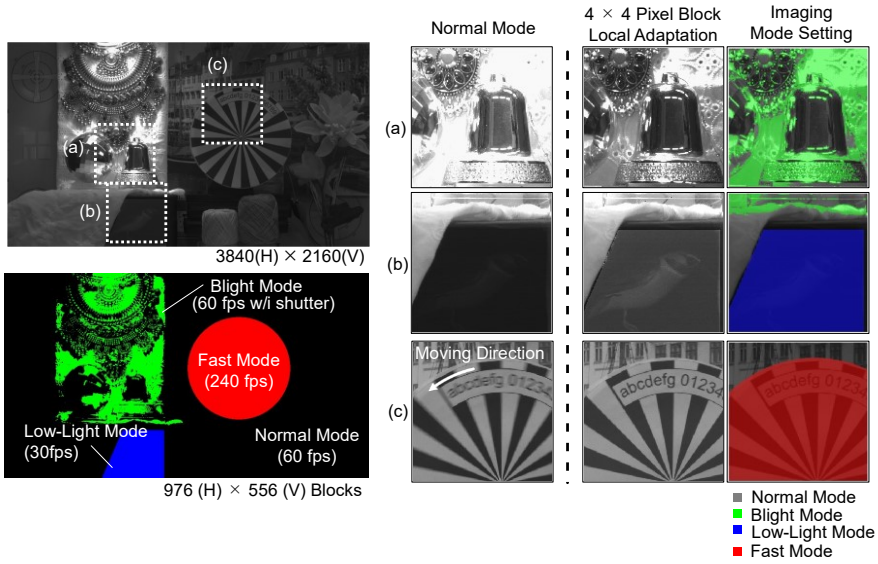


Figure 4: Acquired images with local adaptation for each 4 × 4 pixel block.

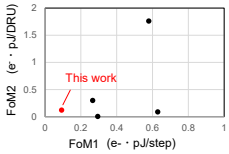
Fabrication Process	BSI 90nm 1P6M
Supply Voltage	3.3 V / 1.2 V
Pixel Size	2.5 μm × 2.5 μm
Pixel Number	3,904 (H) × 2,224 (V)
Block Number	976 (H) × 556 (V), 4 × 4 pixels/block / 488 (H) × 278 (V), 8 × 8 pixels/block
Frame Rate	240 fps (Fast mode) 60 fps (Normal mode) 30 fps (Low-light mode) / 480fps (Fast mode) 120fps (Normal mode) 60fps (Low-light mode)
Feedback Signal Interface	13.5 MHz SPI, 38 data lanes, 1 clock lane
A/D Conversion	14-bit Digital CDS with 6 Times Multiple-Sampling / 12-bit Analog CDS
RMS Random Noise	2.4 e-rms (Gain = 2) / 4.4 e-rms (Gain = 2)
Output Interface	48 data lanes, 4 clock lanes, 576 MHz/lane, SLVS
Conversion Gain	68.2 μV/e-
Full Well Capacity	7,750 e-
Power Consumption	2,640 mW
Die Size	14.2 mm (H) × 17.8 mm (V)

Figure 5: Chip characteristics .

		This work (240fps)	This work (480fps)	[4] OJCS2022	[5] IISW2015	[6] IEEE TCAS2020
Process Technology		**90 nm BSI**		90 nm/40 nm stacked	180 nm	130 nm FSI
Pixel Pitch (μm)		**2.5**		2.7	5	6.5
Number of Pixels (pixels)		**3,904 (H) × 2,224 (V)**		4,224 (H) × 4,224 (V)	800 (H) × 512(V)	256 (H) ×256 (V)
Number of Pixel Blocks (blocks)		**976(H) × 556(V)**		264 (H) × 264 (V)	25 (H) × 16 (V)	256 (H) ×256 (V)
Controllable Resolution (pixels/block)		**4×4**	**8×8**	16×16	32×32	1×1
Controllable parameters	Resolution	**Yes**		No	No	No
	Frame Rate	**Yes**		Yes	No	Yes
	Exposure Time	**Yes**		Yes	Yes	Yes
ADC Resolution (bits)		**14**	**12**	12	10	10
Maximum Frame Rate (fps)		**240**	**480**	1000	100	-
Adaptive Dynamic Range (dB):		**88.2**	**83.0**	110.3	120	74.6

Figure 6: Performance comparison.

	This work	OJCS2022 [5]	ISSCC2022 [8]	VLSI2021 [11]	T-ED2022 [12]	IISW2023 [4]	ISSCC2018 [3]
Process Techonology	90nm BSI	65nm/65nm Stacked	90nm/40nm Stacked	65nm/28nm Stacked	45nm/65nm Stacked	NA	110nm FSI
Number of pixels (M pixels)	8.7	17.8	4.9	2.3	0.3	316.2	33.2
Pixel pitch (μm)	2.5	2.7	1.4	4.95	4.6	4.3	2.1
Shutter operation	RS	RS	RS	GS	GS	RS	RS
ADC bit depth(bits)	14	12	10	10	10	12	12
Frame fate (fps)	240	1000	50	960	30	120	240
Gain (times)	2	1	1	16	1	1	2
Noise (e-)	1.2*	2.9	1.8	2.6	4.2	1.8	4.3
Power (mW)	2640	7400	88	497.8	5.75	23000	9800
Full well capacity (e-)	31000**	947200	13580	14000	9000000	6600	7600
Dynamic range (dB):	88.2	110.3	77.6	74.6	126.6	71.3	64.9
FoM1 (e- · pJ/step)	0.093	0.294	0.629	0.579	2.999	0.266	1.292
FoM2 (e- · pJ/DRU)	0.1177	0.0037	0.0854	1.7608	0.0014	0.2975	5.9886
Architecture	Block control	Block control	Programable resolution	In Pixel ADC	In Pixel ADC	ADC array	3-stage ADC



* Equivalent noise / ** Equivalent FWC
FoM1 = (power × noise)/{Number of pixel × frame rate × $2^{\text{ADC bit depth}}$}
FoM2 = (power × noise)/(Number of pixel × DRU);
DRU = {(saturation/gain/noise}
Dynamic range = 20log(Full well capacity/Noise)

Figure 7: FoM comparison with recent CISs.

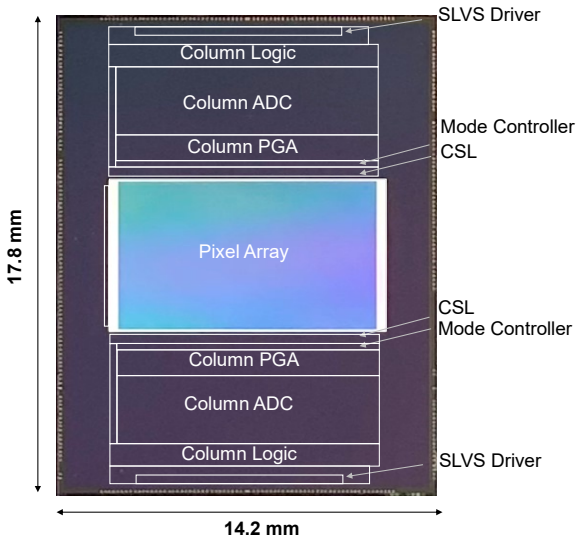


Figure 8: Chip micrograph.

I. REFFERENCES

[1] Recommendation ITU-R BT. 2123: "Video Parameter Values for Advanced Immersive Audio-Visual Systems for Production and International Programme Exchange in Broadcasting," 2019.

[2] S. Kawahito, "Column-Parallel ADCs for CMOS Image Sensors and Their FoM-Based Evaluations," IEICE Trans. Electron., vol. E101-C, no. 7, pp. 444–458, 2018.

[3] Yasue, et al., "A 33Mpixel CMOS imager with multi-functional 3-stage pipeline ADC for 480fps high-speed mode and 120fps low-noise mode," ISSCC Dig. Tech. Papers, pp. 90-92, Feb. 2018.

[4] A. Agarwal, et al., "A 316MP, 120FPS, High Dynamic Range CMOS Image Sensor for Next Generation Immersive Displays," IISW, R7, May 2023.

[5] T. Hirata, et al., "A Vision System With 1-inch 17-Mpixel 1000-fps Block-Controlled Coded-Exposure Stacked-CMOS Image Sensor for Computational Imaging and Adaptive Dynamic Range Control," IEEE Open Journal of Circuits and Systems, Vol. 3, pp. 311-323, Nov. 2022.

[6] A. Peizerat et al., "A 120dB DR and 5μm pixel pitch imager based on local integration time adaptation," IISW, 13. 03, May 2015.

[7] Jie Zhang, et al., "A Closed-Loop, All-Electronic Pixel-Wise Adaptive Imaging System for High Dynamic Range Videography," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 67, no. 6, pp. 1803–1814, May 2020.

[8] H. Murakami, et al., "A 4.9Mpixel Programmable-Resolution Multi-Purpose CMOS Image Sensor for Computer Vision," ISSCC Dig. Tech. Papers, pp. 104–105, Feb. 2022.

[9] K. Tomioka et al., "Feedback Control of a Block -Wise -Controlled Image Sensor Based on Brightness Distribution Analysis," IISW, P25, May 2023.

[10] M. W. Seo, et al., "A Low-Noise High Intrascene Dynamic Range CMOS Image Sensor With a 13 to 19b Variable-Resolution Column-Parallel Folding-Integration/Cyclic ADC," IEEE Journal of Solid-State Circuits, Vol. 47, no. 1, pp. 272-283, Jan. 2012.

[11] M. W. Seo, et al., "A 2.6 e-rms Low-Random-Noise, 116.2 mW Low-Power 2-Mp Global Shutter CMOS Image Sensor with Pixel-Level ADC and In-Pixel Memory," IEEE Symp. VLSI Tech., 2021.

[12] R. Ikeno, et al., "A 4.6-μm, 127-dB Dynamic Range, Ultra-Low Power Stacked Digital Pixel Sensor With Overlapped Triple Quantization," IEEE Transactions on Electron Devices, vol. 69, no. 6, pp. 2943-2950, June, 2022.