A 39,000 Subexposures/s CMOS Image Sensor with Dual-tap Coded-exposure Data-memory Pixel for Adaptive Single-shot Computational Imaging

Rahul Gulve*[1], Navid Sarhangnejad*[1], Gairik Dutta[1], Motasem Sakr[1], Don Nguyen[1], Roberto Rangel[1], Wenzheng Chen[2], Zhengfan Xia[1], Mian Wei[2], Nikita Gusev[1], Esther Y. H. Lin[2], Xiaonong Sun[1], Leo Hanxu[1], Nikola Katic[1], Ameer Abdelhadi[1], Andreas Moshovos[1], Kiriakos N. Kutulakos[2], Roman Genov[1]

1 Dept. of ECE, University of Toronto, Toronto, Canada. 2 Dept. of CS, University of Toronto, Toronto, Canada.

Abstract

A dual-tap coded-exposure-pixel (CEP) image sensor is presented and demonstrated in several computational imaging applications. The NMOS-only data-memory pixel (DMP) reduces transistor count in CEP yielding 39,000 subexposures/s at 320x320 sensor resolution with 7 µm pitch. The outputs of a frame-rate 12-bit ADC1 and a 1-bit subexposure-rate ADC2 are adaptively combined to boost the native dynamic range of coded-imaging modalities by over 57dB, demonstrating over 101dB dynamic range in intensity imaging. The CEP camera combined with machine learnt projection patterns enables single-shot structured-light 3D imaging at native resolution and video rate.

Keywords: CMOS image sensors, high-speed imaging systems, 3D imaging, high dynamic range imaging

Introduction

High-frame-rate image sensors reduce motion artifacts, such as motion blur and ghosting, but may be prone to high read noise, power consumption, and output data rate. The emerging class of coded-exposure pixel (CEP) image sensors [1-3] eliminate these drawbacks by implementing exposure-time programmability for each individual pixel in order to perform multiple subexposures for a single readout, while reducing motion artifacts. This has enabled new applications such as compressive sensing [1,2] and 3D imaging [3], performed in a single shot at native resolution. Other coded-exposure image sensor designs use pixel subarrays that share the same code among many [4,5] or all [6] pixels. As shown in Fig. 1 (left), the total exposure time of one frame is divided into multiple (N) programmable subexposures, or "subframes." The photogenerated charge is selectively accumulated on one or two nodes known as "taps", as controlled by 1-bit binary coefficients, or "codes" organized in frame-sized matrices referred to as "masks." In the case of CEP image sensors, arbitrary codes can be sent to each pixel individually, in each subframe, for fine per-pixel temporal control of exposure. The taps are read out only once per frame resulting in a lower read noise compared to high-frame-rate cameras where a fast readout contributes noise in each short frame. We present a 2-tap CEP image sensor that includes a true single-shot adaptive HDR mode as well as other emerging single-shot imaging modalities, such as auto-tuned structured-light 3D imaging, depth-gated imaging and multispectral imaging.

Architecture and Operation

As shown in Fig. 1 (right), this CEP image sensor includes two ADCs, ADC1 and ADC2, which operate at the frame rate and the subframe rate, respectively. Similar to conventional 2-tap image sensors, ADC1, a 12-bit 2nd-order incremental ΔΣ ADC, provides a high-resolution readout of both taps at a low-speed standard video rate (up to 100fps). ADC2 is a high-speed comparator that compares the taps' voltage with a near-saturation reference voltage in each subframe. In the single-shot HDR application, per-pixel 1-bit ADC2 output from each subframe is used to decide the mask

for next subframe to avoid saturation.

Figure 2 depicts the closed-loop system comprised of two ICs: a 110nm-CIS image sensor and a 65nm-CMOS mask generator and their micrographs. The image sensor shown in Fig. 2 (left) includes a 320x320 array of 2-tap DMP . The two column-parallel ADCs, ADC1 and ADC2, digitize the taps outputs at the maximum frame rate of 100fps, and the peak subframe rate of 39,000sfps (subframes-per-second), respectively. To reduce the power of wireline communication and external memory, the CIS image sensor can be stacked with a digital-CMOS mask generator, such as the one shown in Fig. 2 (right). The mask generator IC includes: (1) a custom low-power mask generator (2) a RISC-V processor and (3) a lossless Huffman-decompression engine each for different types of masks and power requirement. The two dies were tested separately, not in a stack, for the ease of experimental characterization.

The pixel implementation is shown in Fig. 3. All existing CEP image sensor pixels [1-3] belong to the class of Code-Memory Pixels (CMP). They require in-pixel digital memory with PMOS transistors to store the exposure code at the cost of a large and slow pixel. Here, we introduce an NMOS-only two-tap Data-Memory Pixel (DMP) architecture that eliminates the need for in-pixel storage of the exposure code and yields a 7 µm-pixel. Figure 3 (left) shows the equivalent pixel schematic and timing diagram of operation. Figure 3 (right) shows the layout and charge transfer from the photodiode to each tap through an intermediate charge-storage node called "data-memory". The transfer takes 80ns per row or 25.6 µs per subframe for the whole array. The photogenerated charge across all subframes of a frame is selectively integrated on the two taps according to the per-pixel code sequence and is read out once at the end of the frame as two images.

Applications and Results

Figure 4 shows the scene-adaptive single-shot HDR imaging results captured using the combination of ADC1 and ADC2 outputs for N = 213 subframes. Fig. 4 (left) shows the masks at 15 different subframes within one frame exposure period. The mask for the subframe [n] is equal to the output of ADC2 in subframe [n-1]. Fig. 4 (middle-left) shows the per-pixel exposure time realized using the ADC2 output for the adaptive mask control, including three insets with pixels with mostly low (cyan), medium (red) and high (blue) integration time. The HDR image is calculated by dividing ADC1 output with per-pixel exposure time. The tone-mapped HDR image scaled to 8-bits is shown in Fig. 4 (middle-right). Figure 4 (right) depicts the experimentally measured dynamic range of 101.5dB at 30fps, where pixel coding boosts the native (non-coded) pixel dynamic range of 54dB by an additional 57.5dB. Pixel coding also eliminates dips in the signal-to-noise ratio (SNR) for larger signals.

Figure 5 (top, left) depicts the principle of operation of the CEP image sensor in structured-light 3D imaging. Here, 4 optimal illuminations patterns, discovered by stochastic

gradient descent (SGD)[7], are projected onto the scene, and, synchronously, 4 masks with Bayer-like mosaic pattern are submitted to the camera. The sorted photogenerated charge is read out as two images per frame which are used to compute disparity and albedo maps. Figure 5 shows the scene, and the reconstructed 3D maps of the scene captured at native resolution and 30fps video rate using both conventional analytical patterns and the optimized patterns
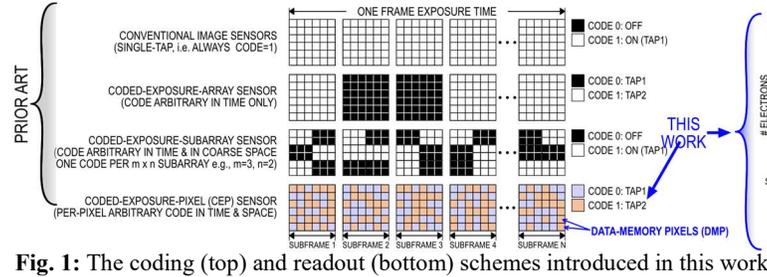
Figure 6 compares the presented CEP image sensor with the best coded-exposure image sensors.

References
(1) N. Sarhangnejad et al., ISSCC, pp. 102-104, Feb. 2019.
(2) Y. Luo et al., Optics Express, pp. 31475-31489, Oct. 2019.
(3) J. Zhang et al., Optics Express, pp. 9013–9024, Apr. 2016.
(4) T. Hirata et al., ISSCC, pp. 120-122, Feb. 2021.
(5) F. Mochizuki et al., ISSCC, pp. 1-3, Feb. 2015.
(6) G. Wan et al., JSSC, pp. 1031-1042, Apr. 2012.
(7) W. Chen et al., CVPR, pp. 5970-5980, Jun. 2020.

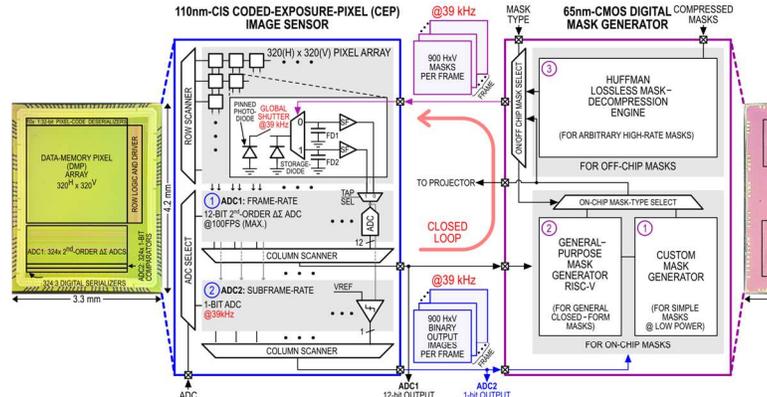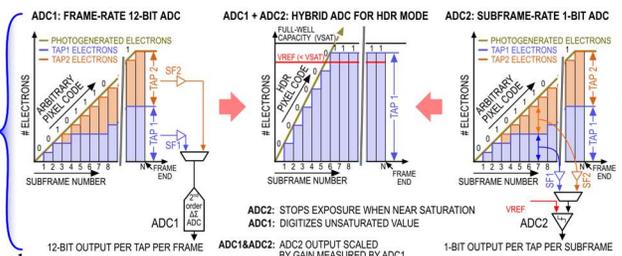**Fig. 1:** The coding (top) and readout (bottom) schemes introduced in this work.

Fig. 2: Closed-loop single-shot coded imaging system consisting of CEP image sensor and digital mask generator, and their micrographs.

Fig. 3: Data-memory pixel (DMP) schematic, timing diagram, layout, and potential well diagram.
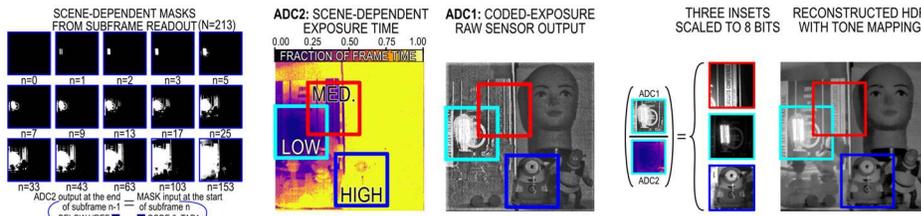

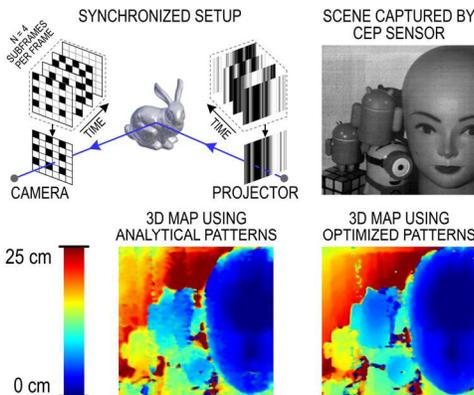**Fig. 4:** Experimentally measured adaptive single-shot HDR imaging results.

Fig. 5: Experimentally measured optimal structured-light 3D imaging results.

BOLD font denotes the best performance among per-pixel coded sensors
<u>UNDERLINE</u> denotes the best performance among all coded sensors
1: $Tap\ Contrast = (Q\_1 - Q\_2)/(Q\_1 + Q\_2)$, where $Q\_{1,2}$ is charge in tap$_{1,2}$
2: FoM=Power/(Number of Pixels×Frame Rate)   3: sfps: subframes per second
4: ROI: arbitrary region of interest   5: only in burst mode

		THIS WORK	[1] Toronto ISSCC 2019	[2] UBC OE 2019	[4] Nikon ISSCC 2021	[5] Shizuoka ISSCC 2015	[6] Stanford JSSC 2012
CODED-EXPOSURE MODE		PER-PIXEL CODING (i.e. PIXELWISE SPATIAL AND TEMPORAL CODING)			PER 16x16 PIXELS	PER 1/15 ARRAY	PER-FULL-ARRAY CODING
PIXEL							
TECHNOLOGY [nm]		**110 CIS / 65 CMOS**	110 CIS	130 CIS	<u>65 CIS / 65 CMOS</u>	110 CIS	130 CIS
PIXEL PITCH [µm]		7 (DMP)	11.2 (CMP)	10.2 (CMP)	<u>2.7</u>	11.2 X 5.6	5
FILL FACTOR [%]		38.5	**45.3**	41.5	<u>100 (BSI)</u>	--	42
NUMBER OF TAPS		2	2	2	1	1	2
TAP CONTRAST [%]¹		96.8 at 39000sfps³	99 at 180sfps³	--	N/A	N/A	--
SYSTEM							
PIXEL COUNT [HxV]		**320 x 320 ≈ 100k**	244 x 162 ≈ 40k	128 x 128 ≈ 16k	<u>4.2k x 4.2k ≈ 17.8M</u>	64 x 108 ≈ 7k	640 x 576 ≈ 369k
MAX READOUT RATE [fps]		ADC1:100 / ADC2:39k	25	10	<u>1000</u>	200M⁵	--
POWER [mW]		107 CIS / 15.2 CMOS	34.4	<u>1.4</u>	--	1620	N/A
POWER FoM [nJ/pixel]²		11	34	<u>8.5</u>	--	N/A⁵	
CODING							
IN-PIXEL DIG. CODE MEMORY (REQUIRES PMOS)		<u>NO</u>	YES (2 LATCHES)	YES (DRAM-LIKE)	NO (STACKED, PER SUBARRAY)	YES (PER SUBARRAY)	N/A
IN-PIXEL ANA. DATA MEMORY		<u>YES (CHARGE)</u>	NO	NO	NO	NO	
SUBFRAME RATE [sfps³]		<u>39k @100k pixels</u>	180 @40k pixels	1.28k @16k pixels	1k @ 69k blocks	200M @ 15 blocks	
CODE RATE [Mbps]		<u>4000</u>	7.1	21	69.7	3000⁶ (128 bits)	
ARBITRARY CODE / ROI⁴		<u>YES/YES</u>	YES/YES	YES/--	NO/--	YES/NO	
FRAME-CODE SHUTTER		<u>GLOBAL</u>	GLOBAL	ROLLING	GLOBAL	GLOBAL	
COMPARISON OF DEMONSTRATED IMAGING APPLICATIONS		SINGLE-SHOT: 1 Adaptive HDR 2 Structured-light 3D (auto-tuned patterns) 3 Depth-gating 4 Direct/indirect imaging 5 Multispectral	SINGLE-SHOT: 1 Structured-light 3D (sub-optimal patterns) 2 Photometric stereo	SINGLE-SHOT: 1 Spatio-temporal compressive sensing, at low resolution and low frame-rate	PSEUDO-SINGLE-SHOT: Adaptive HDR (one-frame latency, on coarse 16x16 pixel subarray grid)	ONE-TIME BURST: Ultra-high-speed recording of short video clip with compressive sensing; not real-time	SINGLE-SHOT: Non-adaptive HDR imaging (with two subframes with two fixed gains), no spatial coding

Fig. 6: Comparison table.