Compact all-CMOS spatiotemporal compressive sensing video camera with pixel-wise coded exposure

Jie Zhang,[1,*] Tao Xiong,[1] Trac Tran,[1] Sang Chin,[1,2,3] and Ralph Etienne-Cummings[1]

[1]*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA*
[2]*Draper Laboratory, Cambridge, MA 02139, USA*
[3]*Department of Computer Science, Boston University, Boston, MA 02215, USA*
[*]*jzhang41@jhu.edu*

Abstract: We present a low power all-CMOS implementation of temporal compressive sensing with pixel-wise coded exposure. This image sensor can increase video pixel resolution and frame rate simultaneously while reducing data readout speed. Compared to previous architectures, this system modulates pixel exposure at the individual photo-diode electronically without external optical components. Thus, the system provides reduction in size and power compare to previous optics based implementations. The prototype image sensor (127 $\times$ 90 pixels) can reconstruct 100 fps videos from coded images sampled at 5 fps. With 20$\times$ reduction in readout speed, our CMOS image sensor only consumes 14μW to provide 100 fps videos.

OCIS codes: (110.1758) Computational imaging; (110.0110) Imaging systems.

References and links

1. G. Bub, M. Tecza, M. Helmes, P. Lee, and P. Kohl, "Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging," Nat. Methods **7**(3), 209–211 (2010).
2. R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered," ACM Trans. Graphics **25**(3), 795–804 (2006).
3. D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).
4. Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *IEEE ICCV* (IEEE, 2011), pp. 287–294.
5. F. Mochizuki, K. Kagawa, S.I. Okihara, M.W. Seo, B. Zhang, T. Takasawa, K. Yasutomi, and S. Kawahito, "Single-shot 200Mfps 5 3-aperture compressive CMOS imager," in *IEEE ISSCC*(IEEE, 2015), pp. 1–3.
6. T. H. Tsai, P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Spectral-temporal compressive imaging," Opt. Lett. **40**(17), 4054–4057 (2015).
7. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," Opt. Lett. **21**(9), 10526–10545 (2015).
8. R. Koller, L. Schmid, N. Matsuda, T. Niederberger, L. Spinoulas, O. Cossairt, G. Schuster, and A. K. Katsaggelos, "High spatio-temporal resolution video with compressed sensing," Opt. Lett. **23**(12), 15992–16007 (2015).
9. D. Liu, J. Gu, Y. Hitomi, M. Gupta, T. Mitsunaga and S. K. Nayar, "Efficient space-time sampling with pixel-wise coded exposure for high-speed imaging," IEEE Trans. Pattern Anal. Mach. Intell. **36**(2), 248–260 (2014).
10. Y. Oike, and A. E. Gamal, "CMOS image sensor with per-column ΣΔ ADC and programmable compressed sensing," IEEE J. Solid-State Circuits **48**(1), 318–328 (2013).
11. T. H. Tsai, X. Yuan, and D. J. Brady, "Spatial light modulator based color polarization imaging," Opt. Express **23**(9), 11912–11926 (2015)

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

12. X. Yuan, T. H. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," IEEE J. Sel. Top. Signal Process. **9**(6), 964–976 (2015).
13. X. Lin, G. Wetzstein, Y. Liu, and Q. Dai, "Dual-coded compressive hyperspectral imaging," Opt. Lett. **39**(7), 2044–2047 (2014).
14. H. Yu, J. Senarathna, B. M. Tyler, N. V. Thakor, and A. P. Pathak, "Miniaturized optical neuroimaging in unrestrained animals," Neuro Image **113**, 397–406 (2015).

1. Introduction

Modern video sensors face some fundamental trade-offs: power consumption vs. frame rate, pixel resolution vs. frame rate, signal to noise ratio (SNR) vs. motion blur. High frame rate imaging leads to fast pixel readout rate with increased power consumption. Given the same power budget, a camera has to sacrifice pixel resolution in order to provide more frames per second. Finally, high frame rate limits pixel exposure time which degrades scene contrast and SNR.

A number of computational imaging methods have been developed to address these trade-offs. Bub et al. implemented a pixel-wise exposure control mechanism using a digital micro-mirror device (DMD) [1]. Using controlled exposure, the system can use a slow frame rate camera with better SNR and dynamic range for high-speed imaging tasks. Raskar et al. proposed the flutter shutter technique to reduce motion blurring [2]. Instead of exposing all the pixels continuously, the flutter shutter method exposes the pixels using a temporally coded shutter. The added pattern improves invertibility of the blur matrix, increasing the ability to de-wrap a blurred image.

Inspired by the theory of Compressed Sensing (CS) [3], a number of CS-based imaging techniques also emerged to improve spatial and temporal resolution of image sensors [4–6]. CS-based sensors use optical frontend to apply a random pixel wise exposure pattern to the focal plane of the sensor. The image sensor then samples the modulated video. These methods compress a spatio-temporal video into a single image. Using inherent sparsity in natural scenes, the video is then recovered from the compressed image using optimization algorithms.

Previous temporal CS imaging systems have demonstrated high image quality at high reconstruction frame rate. But all the previous implementations (both CS based and non-CS based) use optical apparatus to pre-modulate the video scene before the image sensor. Hence, the size and power consumption of the system cannot be further reduced. For example, Llull et al. demonstrated a coded aperture compressive temporal imaging (CACTI) [7]. This prototype uses coded aperture and piezoelectric stage prior to the CCD image sensor. It is able to reconstruct 148 frames per coded image. Koller et al. also presented a prototype compressive video camera at 740 fps using CMOS sensors and silicon-dioxide optical coded mask [8]. Tsai et al. extended this technique to compress a multi-spectral, high-speed scene into a monochrome scene. This system is implemented using objective lens, coded aperture, piezoelectric stage and monochrome CCD camera [6]. It is able to reconstruct 17 spectral channels and 11 temporal channels from a single measurement. Finally, Hitomi et al. proposed an efficient space-time sampling mechanism with pixel-wise coded exposure (PCE) [4, 9]. PCE is developed with constraints suitable for CMOS implementation, but the authors did not demonstrate a CMOS architecture. Instead, the authors verified the concept with a prototype using liquid-crystal-on-silicon (LCOS) device to modulate light prior to the image sensor.

In this paper we propose an all-CMOS architecture to address the disadvantage of previous optical implementations. The chip implements PCE method as described by [4, 9]. Using efficient in-pixel exposure storage, PCE can be completed on-chip without using additional optical apparatus. Additionally, the random coded exposure pattern is generated on-chip using efficient digital pseudo-random generator implemented using linear feedback shift-registers (LFSR). The sensor also implements conventional row scan pixel array readout that is common

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

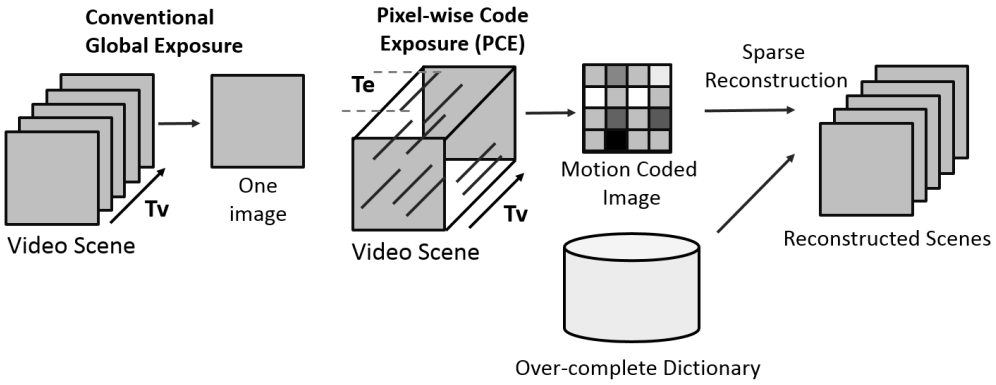

Fig. 1. Conventional Global Exposure imaging vs. Pixel-wise Coded Exposure imaging.

for CMOS image sensors. Thus, the pixel design can be integrated seamlessly into a wide variety of existing systems. Overall, our CMOS architecture reduces the size and power of the previous optical-based systems. This architecture extends the application of pixel-wise coded video compressive sensing to small and low power sensor nodes.

The paper is organized in the following structure: In section 2, we introduce the PCE imaging approach described in [4]. In section 3, we demonstrate a CMOS image sensors architecture that implements PCE using in-pixel memory. In section 4, we demonstrate the measured results from the fabricated chip. In section 5, we discuss the advantages and limitations of the CMOS implementation compared to an optical system. Finally, we conclude the paper in section 6.

2. Pixel-wise coded exposure (PCE) imaging

2.1. PCE vs. conventional imaging

Figure 1 shows the difference between a conventional camera with global exposure and a PCE camera. In a conventional global exposure camera, all the pixels are exposed for a fixed amount of time (Tv) to readout one image at readout frame rate of $1/Tv$. This is compared to a PCE camera, in which pixels are exposed through a random short "single-on" exposure of fixed duration (Te) within Tv. The readout circuit only samples the pixel value at the end of Tv with readout speed of $1/Tv$. PCE essentially compresses a spatiotemporal video into a single coded image. Upon on receiving the coded image, PCE reconstructs the entire video from the single coded image using sparse spatiotemporal reconstruction with an over-complete dictionary. Since the reconstructed framerate is $1/(unit\ time\ of\ Te)$, PCE provides a high frame rate using the same readout speed as a conventional image sensor. PCE is also different from traditional spatial CS approach, which recovers one frame using multiple random spatial samples [5, 10]. Thus, PCE is more optimal for video applications because the sparse samples include both spatial and temporal information. Previous work using optical implementations have shown that PCE is capable of extracting low blur videos from dynamic scenes with occlusions, deforming objects, gas and liquid flow [4].

2.2. Sensing and recovery in PCE

To illustrate sensing and reconstruction steps in PCE mathematically, let there be a spatio-temporal video scene $\mathbf{X} \in \mathbb{R}^{M\times N\times T}$, where $M \times N$ indicates the size of each frame, T indicates the total number of frames in the video and $\mathbf{X}(m,n,t)$ is the pixel value associated with frame t at position (m,n). A sensing cube, $\mathbf{S} \in \mathbb{R}^{M\times N\times T}$, stores exposure control values for pixel

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

at position (m,n,t). The value of $\mathbf{S}(m,n,t)$ is 1 for frames $t \in [t_{start}, t_{end}]$ and 0 otherwise. $[t_{start}, t_{end}]$ denotes the start and end frame numbers for a particular pixel. t_{start} is randomly chosen for every pixel while exposure duration is fixed.

To acquire an image, $\mathbf{Y} \in \mathbb{R}^{M \times N}$, video $\mathbf{X}$ is modulated by the $\mathbf{S}$ before projection across along multiple temporal frames. The value of a pixel $\mathbf{Y}$ at location (m,n) is computed as:

$$\mathbf{Y}(m,n) = \sum_{t=1}^{T} \mathbf{S}(m,n,t) \cdot \mathbf{X}(m,n,t) \tag{1}$$

During reconstruction, we can recover the spatio-temporal video, $\hat{\mathbf{X}} \in \mathbb{R}^{M \times N \times T}$, by solving the following optimization problem:

$$\hat{\mathbf{X}} = argmin_{\mathbf{a}} ||\mathbf{a}||_0 \; s.t. \; ||\mathbf{Y} - \mathbf{SDa}||_2 \leq \varepsilon. \tag{2}$$

where $\mathbf{D} \in \mathbb{R}^{M \times N \times T \times L}$ is the over-complete dictionary learned using training videos. $M \times N \times T$ denote the dimension of the single spatio-temporal dictionary item and L denotes the overall size of the dictionary. $\mathbf{a} \in \mathbb{R}^L$ is the sparse representation of $\mathbf{X}$ using the dictionary $\mathbf{D}$. ε is the tolerable reconstruction error. Similar to previous work, we used K-SVD algorithm for dictionary learning [4, 9].

3. All-CMOS PCE system architecture

3.1. Example of a previous optical PCE system

To construct a PCE camera system, designers must find efficient methods to integrate the video scene with the sensing cube, $\mathbf{S}$, outlined in Eq. (1). Previous work have completed this task using optical apparatus to modulate the light before the image sensor. Figure 2(a) shows an example of the optical system in [4] and [9]. The natural light of the scene first passes through a polarizing beam splitter. This device bounces S-polarized light downward and only allows the P-polarized light to pass through. The S-polarized light is focused on the plane of the liquid crystal on silicon (LCoS) device. When the corresponding LCoS pixel is "on", it alters the light's polarity during the reflection. P-polarized light becomes S-polarized. When LCoS pixel is "off", the polarity of the light stays the same. Only S-polarized light reflects from LCOS can be guided to the image sensor and integrated onto the photodiode according to Eq. (1). The pattern of the sensing matrix, $\mathbf{S}$, is supplied to the LCOS through external control circuits not shown in the diagram.

Although effective at modulating light, optical implementations suffer a few setbacks. First, it is difficult to reduce the size of the system due to additional optical components. Second, extra power is consumed to support the LCOS component. Lastly, the LCOS and the image sensor might have different pixel sizes. Therefore, one-to-one pixel alignment requires careful calibration.

3.2. Electronic pixel-wise exposure control

Figure 2(b) shows our method to implement electrical pixel-wise exposure control directly on the image sensor's focal plane. Each pixel circuit in the image sensor contains a pinned photodiode (*PD*), two switches (*EX* and *RST*) and a buffer to isolate the pixel from the rest of the readout circuits. *PD* can be represented by an equivalent circuit consists of photocurrent, I_{PD}, and diode parasitic capacitance, C_{PD}. I_{PD} is the current through the *PD* as the result of exposure to light. C_{PD} is the parasitic capacitance formed at the PN junction of the diode.

Switch *EX* and *RST* control the start and stop of the pixel's exposure respectively. An example of the pixel operation is shown in the timing diagram of Fig. 2(b). When the signal is set "high", a switch is closed; otherwise it remains open. When the *EX* switch is closed,

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

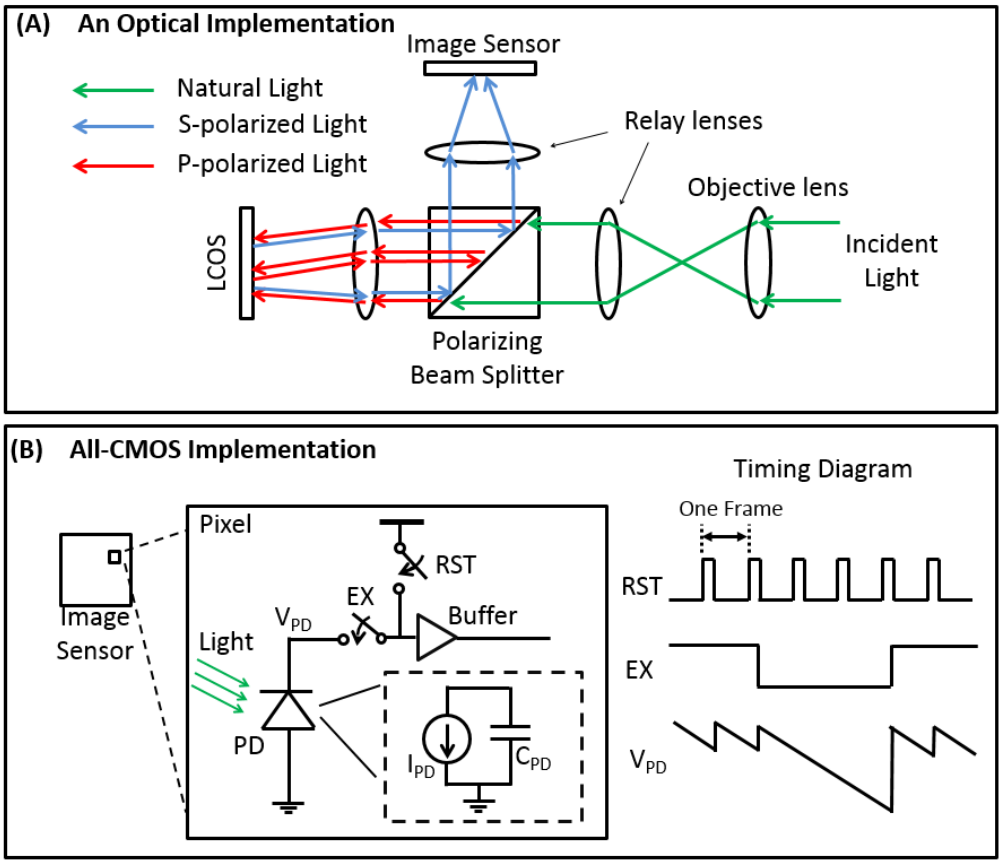

Fig. 2. (A) Optical Implementation of PCE [4, 9] (B) Proposed all-CMOS PCE implementation at the image sensor focal plane.

the toggling of *RST* switch at the beginning of every frame resets the voltage of the *PD*, V_{PD}, to supply rail. On the other hand, when *EX* switch is open, it isolates V_{PD} from the input of the buffer. V_{PD} continues to discharge regardless of the *RST* signal. The pixel ends its exposure when *EX* toggles high. Thus through controlling the duration of *EX*, we can precisely control the start and end time of a pixel's exposure.

3.3. *Pixel with in-pixel exposure memory*

To effectively control the duration of *EX* switch, we inserted a memory element into each pixel. Figure 3(a) illustrates the circuit implementation of Fig. 2(b) using a complementary metal-oxide semiconductor (CMOS) process, with an additional memory element. The pixel consists of two blocks, the Active Pixel Sensor (APS) and the Random Access Memory (RAM). The APS block implements the *EX* and *RST* switches in Fig. 2 with transistor $M1$ and $M3$. $M2$ is the switch used for correlated double sampling (CDS), a mechanism to reduce picel reset noise. $M4 - M6$ forms the buffer and row selection control.

The RAM block is a 1-bit Static RAM (SRAM) which stores the exposure control bit. When *EX* is low, *PD* is separated from the rest of the circuit and continues to discharge its value across multiple frames without reset. When *EX* is high, the photodiode is connected to the rest of the readout circuit. Its value is then read out before reset. To write a new value into the RAM, RAM

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

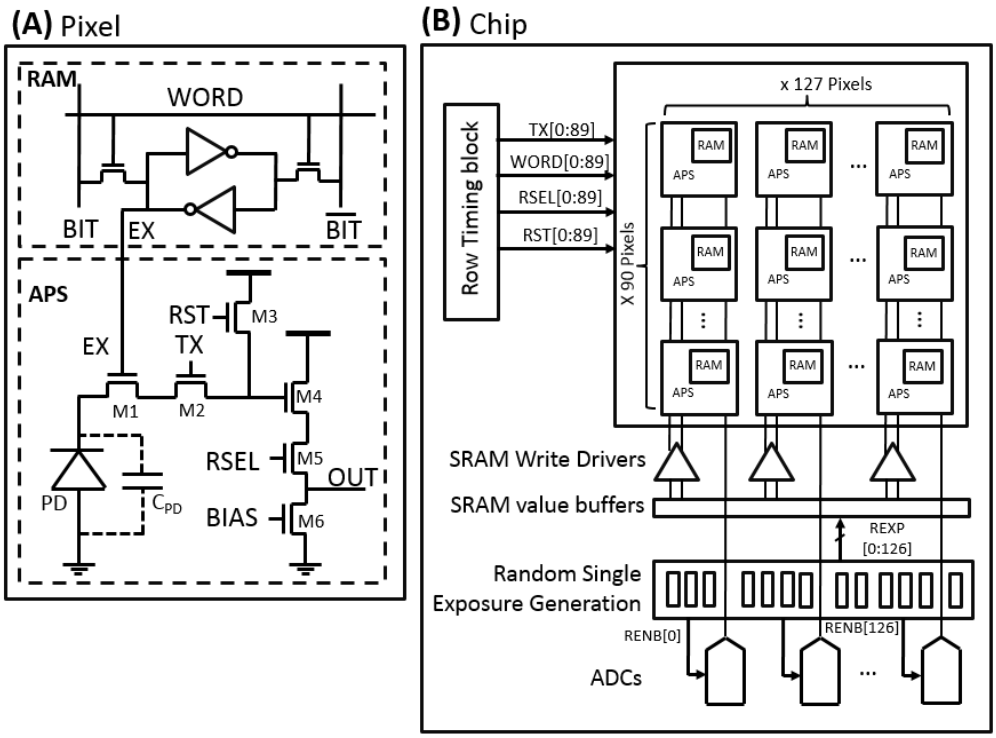

Fig. 3. System architecture for CMOS PCE chip.

drivers placed at the end of the column first write BIT and $\overline{BIT}$. The value are then written into the RAM by toggling the $WORD$ signal.

3.4. Global chip architecture

The system block diagrams for the 127×90 pixel array is shown in Fig. 3(b). Row scan is used to readout the pixel array. Global signal TX, $RSEL$, RST and $WORD$ control the readout timing. To access the in pixel RAM block, SRAM drivers are implemented at each column. Each SRAM driver receives corresponding bit from the signal vector $REXP[0:126]$ that controls the bit to be written.

We use a Random Single-on Exposure Generation (RSEG) block to generate the exposure control bits, $REXP[0:126]$, to be loaded into the in-pixel RAM. RSEG block uses two 7-bit Linear Feedback Shift registers (LFSRs) per row to generate a pseudo random sequence containing the row positions to start and end exposure. Generation of the exposure pattern locally allows higher level of integration and saves power as no additional off-chip memory and circuits are required.

Pixels values are digitized by analog to digital converters (ADC) placed at the end of each column. To save power, ADCs stays idle most of the time except to sample the pixels at the end of their exposure. RSEG also generates control signal vector $RENB[0:126]$ to enable and disable the ADCs.

In this prototype, we demonstrate a sensor with 127×90 pixels. The architecture is scalable to accommodate more pixels. Other than the timing block, number of ADCs, the other system component that scales with the pixel count is the RSEG block. Each row pixels' exposures are

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

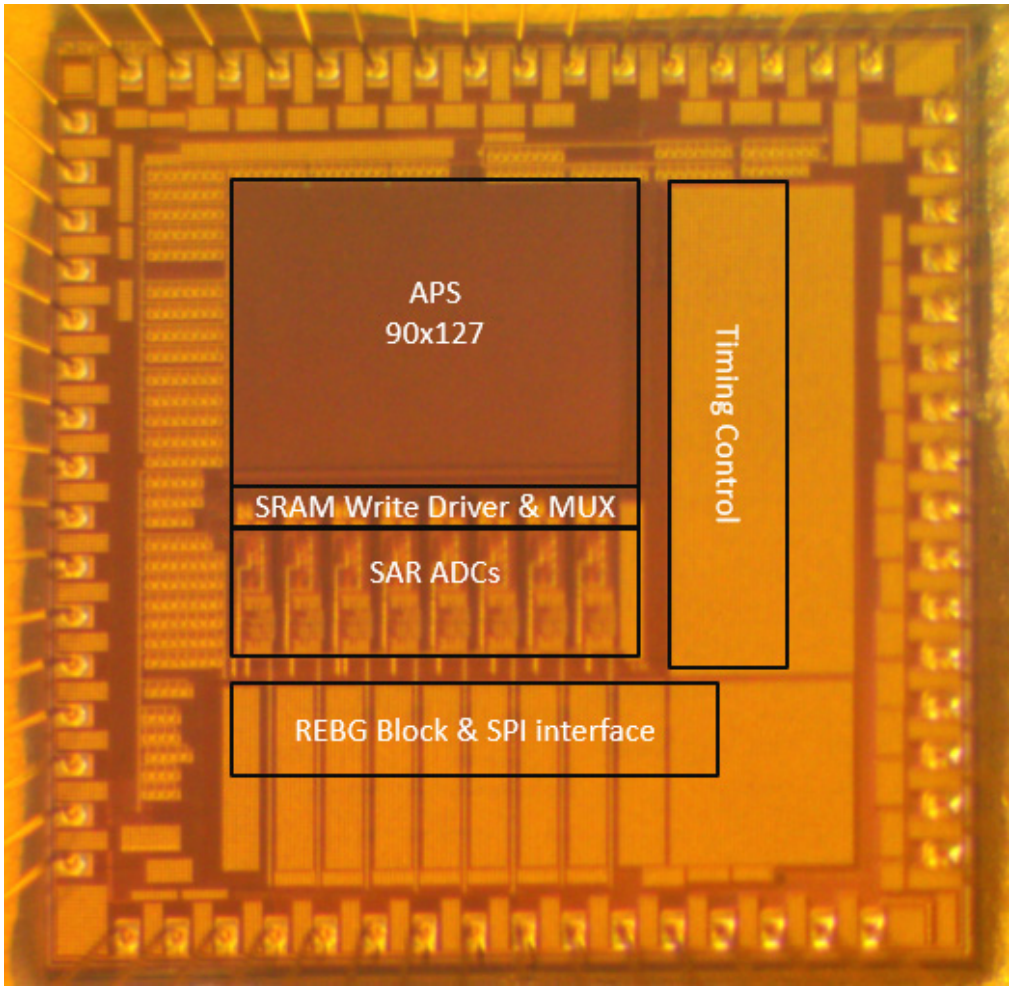

Fig. 4. Chip Micrograph.

controlled by two LFSRs. As we add more rows, more LFSRs need to be inserted. Additionally, As we add more pixels per row, the number of bits in the LFSRs needs to be increased. For example, if pixel count per row increases from 127 to 255, we would need two 8-bits LFSRs instead of the original 7-bits LFSRs. Due to small transistor size for digital circuitry, these registers do not occupy much space on the chip.

4. Sensor measurements

We fabricated the proposed PCE image sensor using a 180nm CMOS process. A Chip micrograph is shown in Fig. 4. The chip occupies an area of $3 \times 3mm$. The pixel array, consist of 127×90 pixels each of dimension $10\mu m \times 10\mu m$, occupies area of $0.9 \times 1.27mm$. Utilizing 6T SRAM structure for exposure storage, the pixel has a fill factor of 52%. Pixel dynamic range ($51.2dB$), fixed pattern noise (1.02%) and random noise (5.4 Digital Number) are limited by the fabrication process. We demonstrate the chip measurement result in this section.

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

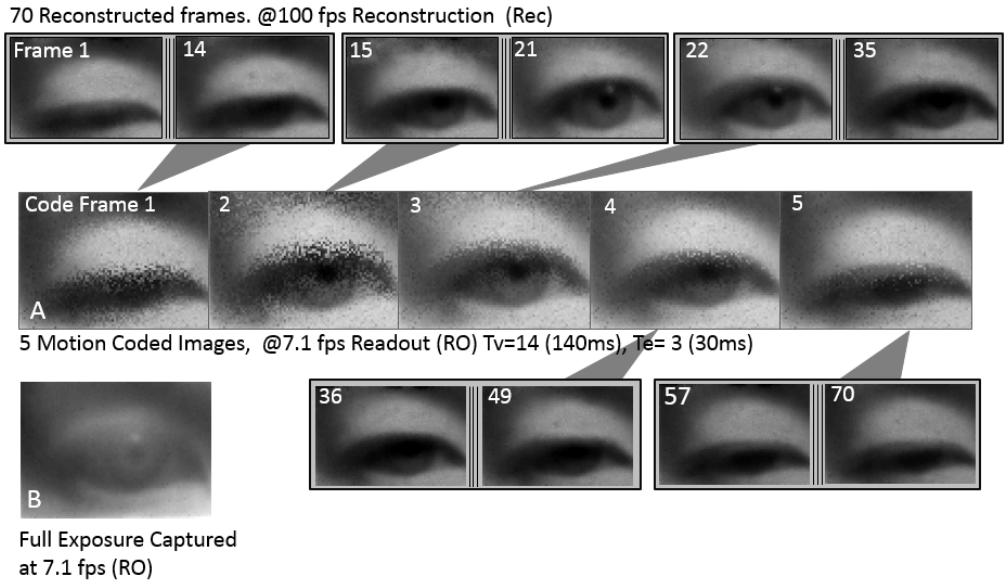

Fig. 5. Reconstructed video scenes using coded exposure images from the sensor.

4.1. Reconstruction of high frame rate video using low frame rate readout

Figure 5 shows a video of a blinking eye recorded by the image sensor. In this example, we set the unit time of Te and Tv to be $10ms$. This means without temporal compression, the image sensor can put out videos at rate of 100 fps. The sensor then compresses 14 frames ($Tv = 14(140ms)$) into one single image through coded exposure. Hence the readout (RO) speed of the sensor is reduced to around 7.1fps (100/14). Each pixel undergoes $30ms$ of exposure, corresponding to $Te = 3(30ms)$.

Five coded images are shown in Fig. 5. From each of these coded images, 14 video frames are reconstructed by solving an L1-optimization problem in Eq. (2). The result of the recovery is a 100 fps equivalent video. As a comparison, at the same RO speed of 7.1 fps, a global exposure captures a severely blurred image, shown in Fig. 5(b).

Video reconstruction is done block-wise. The coded image is broken down into blocks of size 8×8. We then reconstruct a spatio-temporal cube of $8 \times 8 \times 14$ using a dictionary of size 896×30000. To acquire training data for the dictionary, we collected video of various objects and movements at 100 fps frame rate and used K-SVD algorithm to train the over-complete dictionary, similar to previous work [4] and [9].

4.2. Improving SNR and reducing motion blurring simultaneously

Figure 6 shows an example that PCE image sensor can be used to reduce readout speed, enhance video SNR and reduce motion blur at the same time.

ROW1 shows a video of a blinking eye captured by the image sensor at full rate without pixel wise coded exposure. This is a 100 fps video with $10ms$ frame exposure time between frames. The SNR and contrast of the scene is low as the signal level is weak.

On the other extreme, ROW2 shows a video captured with 20 fps with $50ms$ exposure time between frames. The scene SNR increases but at the cost of motion blur and reduced framerate.

Our PCE image sensor is then used to capture a similar scene using $Te = 5$, which allows $50ms$ of pixel exposure time. The coded image and recovered frames are shown in ROW3. Due

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

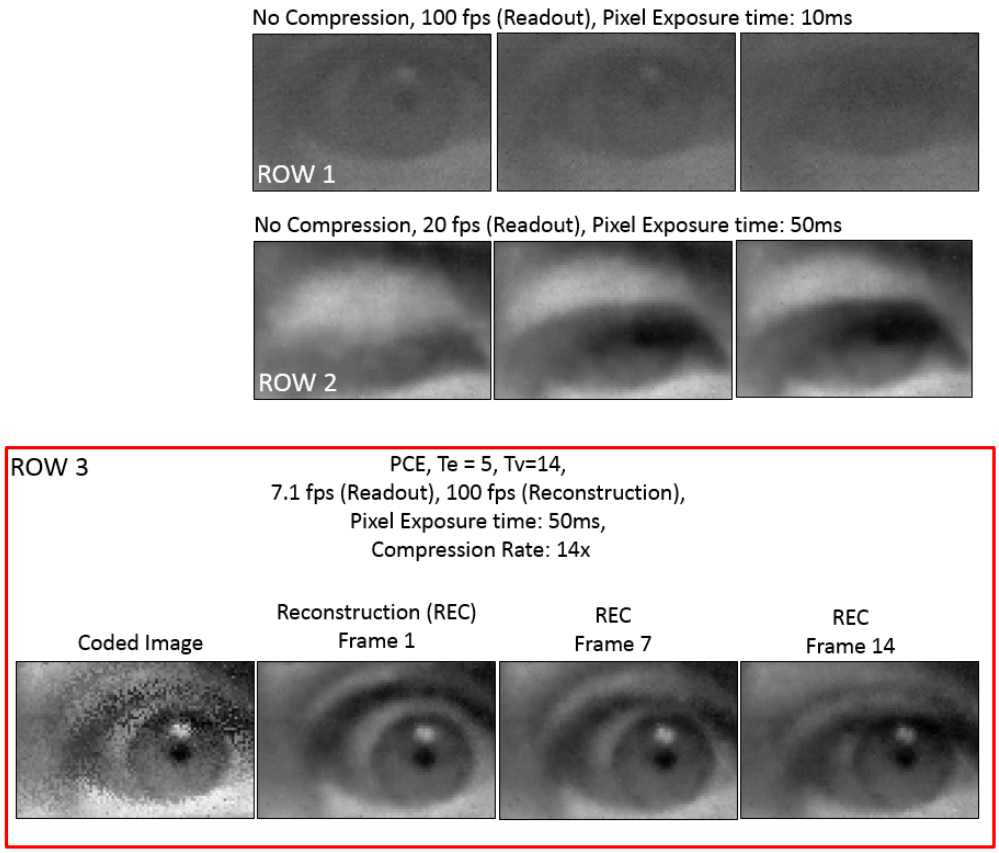

Fig. 6. Enhancing scene SNR and reducing motion blur simultaneously using PCE.

to extra exposure time, SNR of the coded image improves. Since an overcomplete dictionary trained using blur-free 100 fps videos is used for sparse reconstruction, the reconstructed video's blurring is reduced. In comparison, video in ROW3 has higher SNR than ROW1 with lower motion blur compared to ROW2. ROW3's readout rate is also 14$\times$ less than ROW1.

4.3. Power consumption

Figure 7 shows the chip power consumption at different compression rates (CR). For the same Te, longer Tv leads to larger CR and lower power. However, longer Tv also causes the sensor to collect fewer spatial temporal samples per frame. This may lead to degradation of the reconstruction image quality.

The graph in Fig. 7 shows the chip power consumption breakdown at different Tv when Te is fixed ($Te = 3$). When the scene consists of many smooth surfaces, longer Tv can be used to save power while guaranteeing acceptable video reconstruction (REC) quality. A reconstructed example of such scene is shown at $Tv = 20$, where the scene consists of fingers and the smooth surfaces of a chip package. On the other hand, when higher detail of the scene is desired, short Tv can be used to refine the REC quality. An example is shown at $Tv = 5$, where the fine detail of the spikes in the package is well reconstructed. For visual references, lossless images collected at full rate are also shown in the figure. At CR of 20, the chip power consumption is only $14\mu W$, compared to $1.3mW$ at full rate. This corresponds to around 99% of power saving.

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

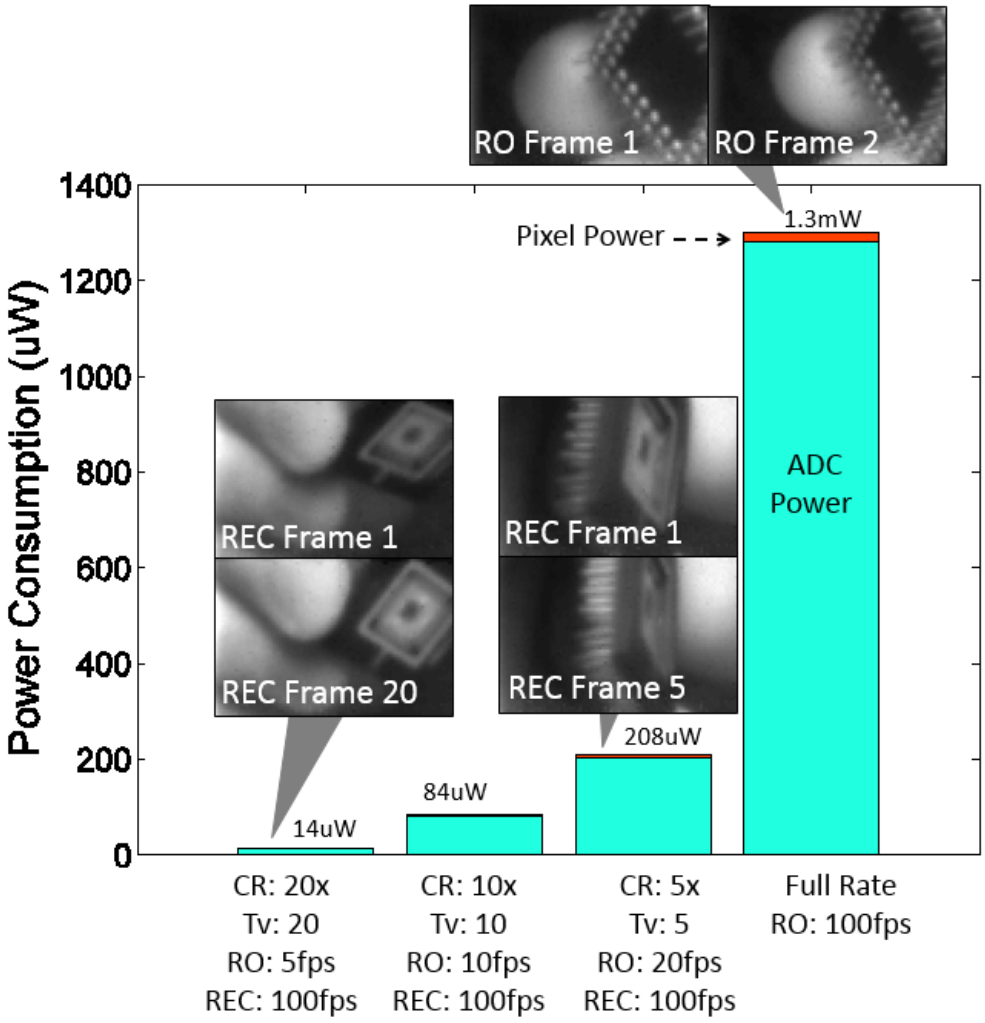

Fig. 7. Power Consumption and Image Quality at different Compression Rate (CR).

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

5. Comparison with previous optical implementation

In this paper we demonstrate a functional CMOS imager with in-pixel memory to implement pixel-wise coded exposure imaging. This architecture offers an alternative to previously demonstrated optical implementations. Due to different system settings (image sensor process, type, resolution, frame rate and etc), a quantitative comparison with the optical counterparts does not yield conclusive results. Nonetheless, we formulate a qualitative discussion in this section to address the advantages and limitations of the proposed CMOS architecture.

5.1. Size

The CMOS implementation is much smaller in size compared to its optical counterparts. Without using additional optical modulators, the entire pixel-wise exposure coding can be done directly on an image sensor with dimension of a few mm^2.

Furthermore, One of the disadvantages of LCoS implementation is the need for polarization which reduces the light throughput. Polarization in the LCoS can also be sensitive to color. There are several methods to apply broad light to LCoS device, but a cost of extra modulation and computation [11–13]. On the other hand, similar to the DMD modulation, all-CMOS implementation does not require light polarization and allows broad band incident light. The photodiode used in this prototype only measures the grayscale level. But conventional pixel filters and RGB pixels can be easily incorporated into this architecture to support color imaging.

5.2. Power consumption

Our CMOS implementation also offers power saving compared to optical counterparts. The CMOS architecture utilizes SRAM block as in-pixel storage. SRAM are efficient storage units as power is only consumed during write operations. It does not consume static power when the bits are stored. On the other hand, optical modulators are often very power consuming. For example, the LCOS device used in [4], ForthDD SXGA-3DM (1280 x 1024 pixels), consumes several Watts of power during full switching operation.

5.3. Speed/frame rate

The speed is comparable for both implementations. The speed of both system is limited by the readout electronics of the image sensor. For the CMOS architecture, since SRAM is capable of operating at several hundred of MHz, the row update of the in-pixel RAM is not a limiting factor to the overall speed of the system. For the optical system, the light modulators can also switch at speed much faster than the readout speed of the image sensor and therefore does not limit the system frame rate. For example, the LCOS used in [4] supports refresh rate up to 3.2KHz. More efficient sliding aperture masks can also be used to modulate the scene at a much faster rate.

5.4. Exposure pattern flexibility

As mentioned in [4], due to the inability to store charge within the pixel photodiode, the CMOS implementation only supports single-on coded exposure patterns. On the other hand, the optical imaging system has less limitation in the choice of random exposure coded masks.

6. Conclusion

We described a CMOS implementation of the pixel-wise coded exposure (PCE) imaging using in-pixel memory. Compared to its optical counter-part, the CMOS implementation is small and power efficient. The 127×90 pixel array is capable of providing video at 100fps while using readout speed of only 5 fps. With $20\times$ compression, the system consumes only $14\mu W$ of power.

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016

This architecture extends the benefits of pixel-wise exposure control cameras to small image sensors, enabling new generation of small and power efficient high frame rate sensors.

Some potential applications for this architecture are miniature image sensors for mobile wireless sensor node (MWSN) and microscopy in unrestrained animals [14]. These applications require high frame rate and low blur image sensors with high performance under different light illumination. The sensor must also satisfy these requirements with low power consumption due to battery limitation in the mobile setting. Our architecture is able to satisfy these specifications by acquiring high frame, high SNR, and low motion blur images at a small fraction of the power compared to a conventional image sensor.

#259324 Received 11 Feb 2016; revised 9 Apr 2016; accepted 9 Apr 2016; published 15 Apr 2016