CMOS IMAGE SENSORS WITH MULTI-BUCKET PIXELS FOR COMPUTATIONAL PHOTOGRAPHY

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF

ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Chung Chun Wan

November 2011

This dissertation is online at: http://purl.stanford.edu/vw150rs2182

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Mark Horowitz, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Marc Levoy, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Philip Wong

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

When applied to multi-image computational photography such as flash/no-flash imaging, multiple exposure high dynamic range imaging, multi-flash imaging for depth edge detection, color imaging using active illumination, and flash matting, an image sensor that can capture multiple time-interleaved images would provide a dramatic advantage over capturing and combining a burst of images having different camera settings. In particular, this interleaving eliminates the need to align the frames after capture. Moreover, all frames have the same handshake or object motion blur, and moving objects are in the same position in all frames. A sensor with multi-bucket analog memories in each pixel can accomplish this task. Whereas frames are acquired sequentially in a conventional sensor, in a multi-bucket sensor photo-generated charges in a photodiode can be transferred and accumulated in the in-pixel memories in any chosen time sequence during an exposure so multiple frames can be acquired virtually simultaneously.

Designing a multi-bucket pixel which is compact and scalable is challenging because space is required to accommodate the additional in-pixel memories and their

associated control signal lines. This research explored and developed a new multi-bucket pixel technology by adapting the concept of virtual phase charge-coupled device into a standard 4-transistor CMOS pixel such that area overhead is small and true correlated double sampling is preserved to cancel kTC noise.

Based on the developed pixel technology, two prototype CMOS image sensors with dual and quad-bucket pixels were designed and fabricated. The dual-bucket sensor consists of 640_Hx576_V 5.0μm pixels in 0.11μm CMOS technology while the next-generation quad-bucket sensor comprises 640_Hx512_V array of 5.6μm pixels in 0.13μm CMOS technology. Pixel sizes are the smallest among similar pixels reported in the literature. Some computational photography applications were implemented using the two multi-bucket sensors to demonstrate their values in avoiding artifacts that would otherwise occur when a conventional sensor is used.

Acknowledgements

First, I would like to express my most sincere gratitude to Prof. Mark Horowitz for being such an incredible advisor. His depth and breadth of technical and non-technical knowledge have impressed me since my very first day at Stanford when I sat in his EE271 class. If an advisor was an image, he achieved the photon shot noise limit. I am very fortunate to have him as my advisor.

Next, I would like to thank Prof. Marc Levoy for being my associate advisor and providing directions and guidances to the computational photography part of my thesis. He is one of the few professors who would sit down and brainstorm ideas with you at his office for hours. Prof. Philip Wong has been a very special person to me in my graduate study at Stanford. Without the opportunity that he had given me, getting a PhD from Stanford would simply be something that I would occasionally dream about. I would also wish to thank Prof. Ada Poon for being the most patient counselor to me whenever I desperately needed one.

This dissertation would not have been possible without the support of my colleagues at Aptina Imaging. Particularly, I would like to thank Dr. Gennadiy Agranov for his endless encouragement and support of this project. I would also like to

specially thank Dr. Xiangli Li for being the most considerate manager I have ever met. I also wish to thank Dr. Jerry Hynecek and Hirofumi Komori for being two of the most important persons from whom I learned about pixel design.

At Stanford, I am very fortunate to have the opportunity to interact with many great members in several research groups, including the Horowitz group, the Levoy group, and the Wong group. I would like to particularly thank some of my good friends at Stanford such as Nishant Patil, Frances Lau, Jenny Hu, Helen Chen, Prasanthi Relangi, Metha Jeeradit, Jie Deng, Gael Close, and Byong Chan Lim.

Furthermore, much gratitude goes to my sisters, my brothers-in-law, and my lovely nephew and nieces.

I wish to thank my fiancée for sharing my ups and downs throughout the years and promising to be with me for the rest of my life. She is the most precious gift that I have got at Stanford.

Finally, I would like to dedicate this dissertation to my heavenly Father and parents for their endless and unconditional love.

"We love, because He first loved us." 1 John 4:19

A picture taken by the quad-bucket sensor described in Chapter 4 of this thesis.

Contents

List of Tables

List of Figures

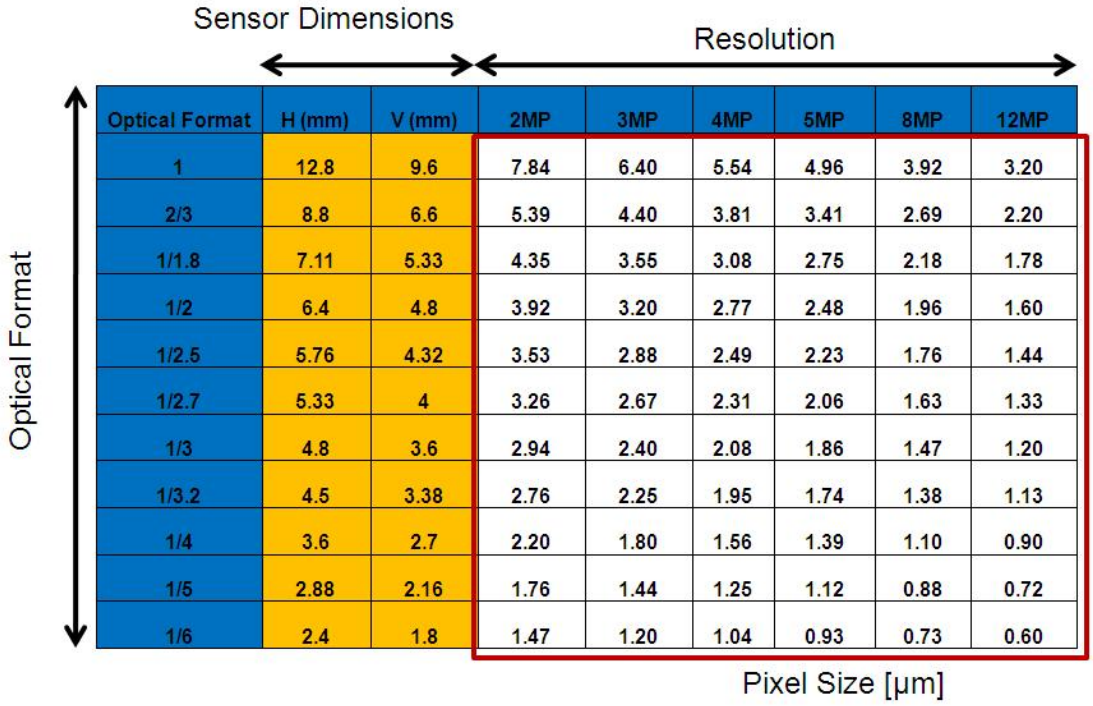
Chapter 1

Introduction

Advancement in digital imaging, in particular image sensor technology, has revolutionized our lives in the last few decades. In order to mimic the best film, the goals of an image sensor have always been to achieve the most pixels, the smallest cost, the highest sensitivity, and the largest signal-to-noise ratio. Despite that many of these goals compete against each other, image sensors today are amazing in terms of pixel count, sensitivity, low-cost, and read noise. Pixel size scaling has so far been the strongest driving force in achieving these goals [Agranov 09] [Agranov 11].

1.1 Recent Pixel Scaling

The initial reason for pixel scaling was increasing pixel count and therefore image quality. However, after pixel count has reached beyond 6 megapixels, further reduction in pixel size is mainly because of marketing concern and form factor reduction.



Optical Format	H (mm)	V (mm)	2MP	3MP	4MP	5MP	8MP	12MP
1	12.8	9.6	7.84	6.40	5.54	4.96	3.92	3.20
2/3	8.8	6.6	5.39	4.40	3.81	3.41	2.69	2.20
1/1.8	7.11	5.33	4.35	3.55	3.08	2.75	2.18	1.78
1/2	6.4	4.8	3.92	3.20	2.77	2.48	1.96	1.60
1/2.5	5.76	4.32	3.53	2.88	2.49	2.23	1.76	1.44
1/2.7	5.33	4	3.26	2.67	2.31	2.06	1.63	1.33
1/3	4.8	3.6	2.94	2.40	2.08	1.86	1.47	1.20
1/3.2	4.5	3.38	2.76	2.25	1.95	1.74	1.38	1.13
1/4	3.6	2.7	2.20	1.80	1.56	1.39	1.10	0.90
1/5	2.88	2.16	1.76	1.44	1.25	1.12	0.88	0.72
1/6	2.4	1.8	1.47	1.20	1.04	0.93	0.73	0.60

Figure 1.1: *The relationship between a sensor's optical format, dimensions, spatial resolution, and pixel size.*

Reducing the form factor is important because it allows shortening the optical path and therefore enabling, for instance a thinner phone. Given that smartphone market has become the biggest for image sensor companies and that a slimmer phone is in general more appealing, form factor reduction has become the highest priority for image sensor manufacturers and reducing pixel size is so far the only way to achieve this goal. Figure 1.1 shows the relationship between optical format (or form factor), sensor dimensions, spatial resolution, and pixel size. As we can see, reducing form factor or increasing spatial resolution both require a smaller pixel.

However, as pixel size gets smaller, pixel performance also goes down. Figure 1.2 shows in particular that the full well capacity and sensitivity drop when a pixel

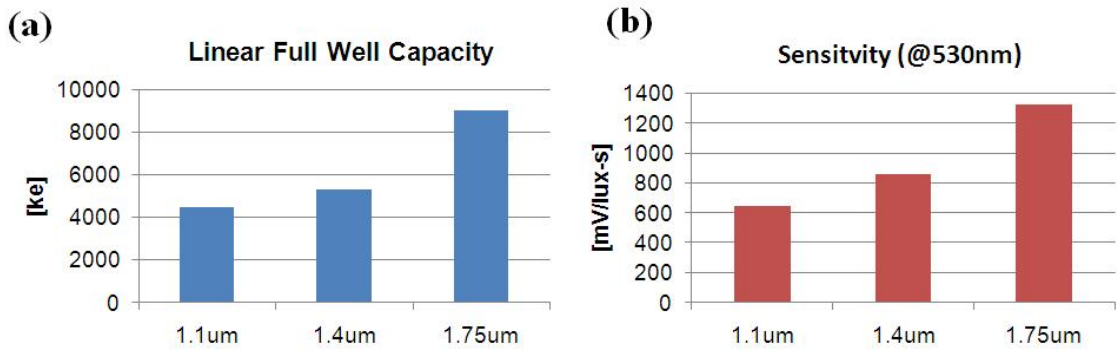


Figure 1.2: *Performance of pixels with state-of-the-art pixel sizes. (a) Full well capacity (b) Sensitivity.*

becomes smaller. The consequence is that image quality will eventually drop to an unacceptable level if pixels keep shrinking. As a result, the pixel scaling will sooner or later slow down. To further image sensor development, alternative directions need to be explored. One of the possible directions is to add new functionalities to an image sensor. But what new functionalities should be added? To answer this question, we digress from image sensors and see what else is happening in imaging technology.

1.2 Computational Photography

During the image sensor improvement, a new approach to create images called computational photography has emerged [Raskar 06a] [Nayar 06] [Hayes 08]. In this approach, a picture is no longer taken but rather computed from image sensor data. This added computation is performed to improve image quality or produce pictures that could not have been taken by traditional cameras. Some representative computational photography techniques include multiple exposure high dynamic range

[Reinhard 06], flash/no-flash [Petschnigg 04] multi-flash [Raskar 04], and panorama. A common motif to a number of computational photography techniques takes the form: "Capture a burst of images varying camera setting X and combine them to produce a single image that exhibits better Y [Levoy 10]." Nevertheless, only a few photographic situations currently lend themselves to this solution because undesired changes other than setting X, such as motion and varying ambient lighting conditions, usually occur in real scenes - people walk, breezes blow, foliage moves, and hands shake [Levoy 10]. The subsequent reconstruction algorithms used to combine the images need to cope with those undesired changes or otherwise the final computed image would be severely degraded. However, it is currently challenging to come up with robust algorithms that can handle all situations [Szeliski 10].

An ideal way to solve the problem seems to be capturing the multiple images simultaneously. However, achieving such simultaneity is impossible in a single camera system that uses a conventional image sensor because the sensor can only take one picture at a time: frames must be acquired sequentially. Depending on the motion and the frame rate of the sensor, the captured images can be significantly different from each other causing artifacts in the final computed picture. On the other hand, capturing multiple images simultaneously is not compatible with some multi-image computational photography techniques such as flash/no-flash imaging in which the scene is being modified sequentially. This thesis discusses a new imaging approach called time-multiplexed exposure in which multiple frames can be captured in a tightly interleaved manner. This interleaving equalizes the undesired changes

in the scene among the captured frames except those synchronized changes that the users intentionally add to the scene (e.g. a flash). The consequence is that the subsequent reconstruction algorithms can be made simple enough to make multi-image computational photography more robust to be used in real scenes.

1.3 Contributions

This thesis makes the following contributions [Wan 11a] [Wan 11b]:

1. Defined an alternative imaging approach called time-multiplexed exposure that enhances and enables multi-image computational photography.

2. Developed a new multi-bucket pixel technology which forms the core of multi-bucket CMOS image sensors that support time-multiplexed exposure.

3. Designed and fabricated two prototype multi-bucket CMOS image sensors with smallest pixel sizes among similar pixels reported in the literature.

4. Demonstrated enhanced and newly enabled computational photography applications using the new image sensors.

1.4 Outline

The thesis is structured as follows:

Chapter 2 introduces time-multiplexed exposure, an alternative imaging approach that can address the current challenges faced by many computational photography algorithms. Two different implementations of this imaging approach are identified

and we discuss the reason why an image sensor with pixel-level analog memories or buckets is a better architecture.

Chapter 3 first reviews the basics of CMOS imager pixel technology to provide a basis for discussing desirable features a multi-bucket pixel should have. Then, we describe a new device structure that uses the concept of virtual phase CCD and forms the basis of our CMOS multi-bucket pixels. The device structure, its operation, and simulation results are presented.

Chapter 4 describes the design, operation, and characterization of the two prototype multi-bucket CMOS image sensors developed in this research. The first sensor has dual in-pixel memory and comprises a 640_Hx576_V array of 5.0μm pixels. The second sensor contains an array of 640_Hx512_V 5.6μm pixels. The first sensor is fabricated in a 2-poly, 4-metal 0.11μm CMOS process while the second one is fabricated in a 2-poly, 3-metal 0.13μm CMOS process.

In Chapter 5, we present applications of our multi-bucket sensors in computational photography. We demonstrate the values of the sensors in avoiding artifacts that would normally be observed when a conventional sensor is used.

Chapter 6 concludes the thesis with a summary of results and a discussion about future direction of this research.

Chapter 2

Time-Multiplexed Exposure

In multi-image computational photography, a burst of images exposed under different camera settings are captured. In a single-camera system that uses a conventional image sensor, the burst of images are captured sequentially as shown in Figure 2.1. The captured images are subsequently combined to create a final image which is superior in some aspects compared with any of the captured images. Representative examples include multiple exposure high dynamic range (HDR) [Reinhard 06], flash/no-flash [Petschnigg 04], multi-flash [Raskar 04], and color imaging using active illumination [Ohta 07].

Generally speaking, the above imaging approach is clever and works nicely in a static scene. However, it is challenging to use this approach in a dynamic scene because undesired changes occur both within (intra-frame) and between (inter-frame) the multiple images. Among the undesired changes, motion (both camera shake and object motion) and varying lighting conditions are the most naturally found

phenomena in real scenes. Depending on the scene, the exposure time of each frame, and the frame rate of the sensor being used, the multiple images captured can be different in an undesired way by varying degrees. For example, a moving object may appear at different positions in the captured images or the captured images have different amount of handshaking blur.

Figure 2.2 shows ghosting artifact due to motion when the objects in the scene move during the burst of images captured for a multiple exposure HDR photograph. In this approach, two or more images with varying exposures are captured and subsequently combined to synthesize a HDR image [Reinhard 06]. The undesired differences caused by, for example motion, in the multiple images usually cause the subsequent reconstruction algorithms to fail and therefore artifacts to appear in the final computed picture - defeating computational photography's purpose of improving image quality.

Many reconstruction algorithms have been devised over the years to avoid the artifacts mentioned above [Gallo 09] [Ward 2003] [Kang 03] [Eden 06] [Mills 09]. Particularly, sophisticated image registrations or alignments are performed before blending the images. While reasonable successes have been demonstrated, the effectiveness of the algorithms is scene and sensor dependent and will not be robust all the time [Szeliski 10]. Especially when the images have widely different exposures, the task of aligning them becomes even more challenging [Eisemann 04]. In fact, all the algorithms have been developed under the implicit assumption that the underlying image capture and hardware stay the same. But, why should it be the case? In this

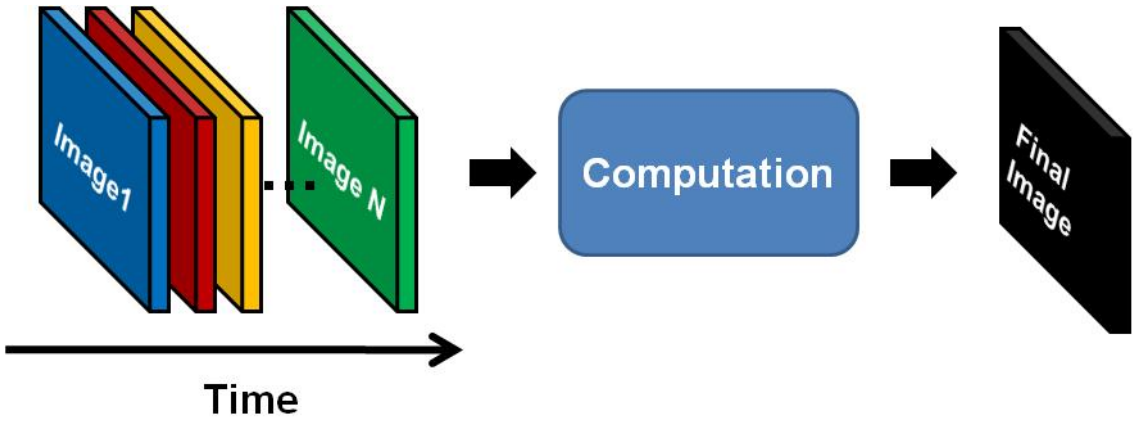


Figure 2.1: *Multi-image computational photography using a conventional image sensor.*

chapter, we relax this assumption and explore how the above problem can be tackled if image capture and hardware are also taken into consideration.

2.1 Limitation of Sequential Image Capture

Conventional image sensors capture images sequentially and their frame rates determine how fast images can be taken successively. Figure 2.3 shows an example of how three images with different exposure times (e.g. in HDR photography) are captured by conventional rolling shutter sensors with different frame rates. Figure 2.3 (a) illustrates the point that even though exposure times of all frames are shorter than the frame time, the next frame cannot start immediately due to the limited frame rate of the sensor. For example, taking three images with 1/125s, 1/250s, and 1/500s back-to-back would require an image sensor to have a frame rate of 500fps. Due to the inter-frame time gaps, moving objects in the scene can appear at different positions in the captured frames.

Figure 2.2: *Ghosting artifact due to motion in multiple exposure HDR photography. The HDR photo was taken by a commercial smart phone inside which an image sensor runs at a maximum of 15fps. Two frames, one long and another short, were taken by the phone to synthesize the HDR photo. A time gap of roughly 1/15s exists between the two frames due to the limited frame rate of the sensor.*

Inter-frame time gaps, together with rolling shutter artifacts, can in principle be improved by increasing the frame rate of the sensor (Figure 2.3 (b) and (c)). However, despite the fact that frame rates of image sensors have been improving steadily over times, practical constraints such as circuit speed and power consumption limit the maximum achievable frame rate especially for sensors with high resolutions.

In fact, even in the ideal situation where a sensor has an infinite frame rate, the captured frames could still be fundamentally different from each other because their exposures start and end at different times. The amount of difference depends on the relative exposure times of the frames and the undesired changes (e.g. how fast an object is moving) in the scene. Figure 2.4 shows sample images captured using sensors with different frame rates. Even though increasing frame rate helps minimizing the inter-frame differences, the intra-frame differences remain.

Due to the fundamental limitation imposed by conventional image sensors, multi-image computational photography typically needs to carefully post-process the captured frames (e.g. image alignment or motion compensation) before combining them to produce a final image. This procedure, as previously mentioned, can be error-prone and create artifacts in the final image as shown in Figure 2.2. The example shown in Fig 2.4 represents an even more challenging situation because blurred and non-blurred images need to be aligned [Yuan 07].

Can we modify our image capture to reduce the needed post-processing or eliminate them altogether to make multi-image computational photography more robust? In fact, given that capturing multiple images sequentially is a concept that has been

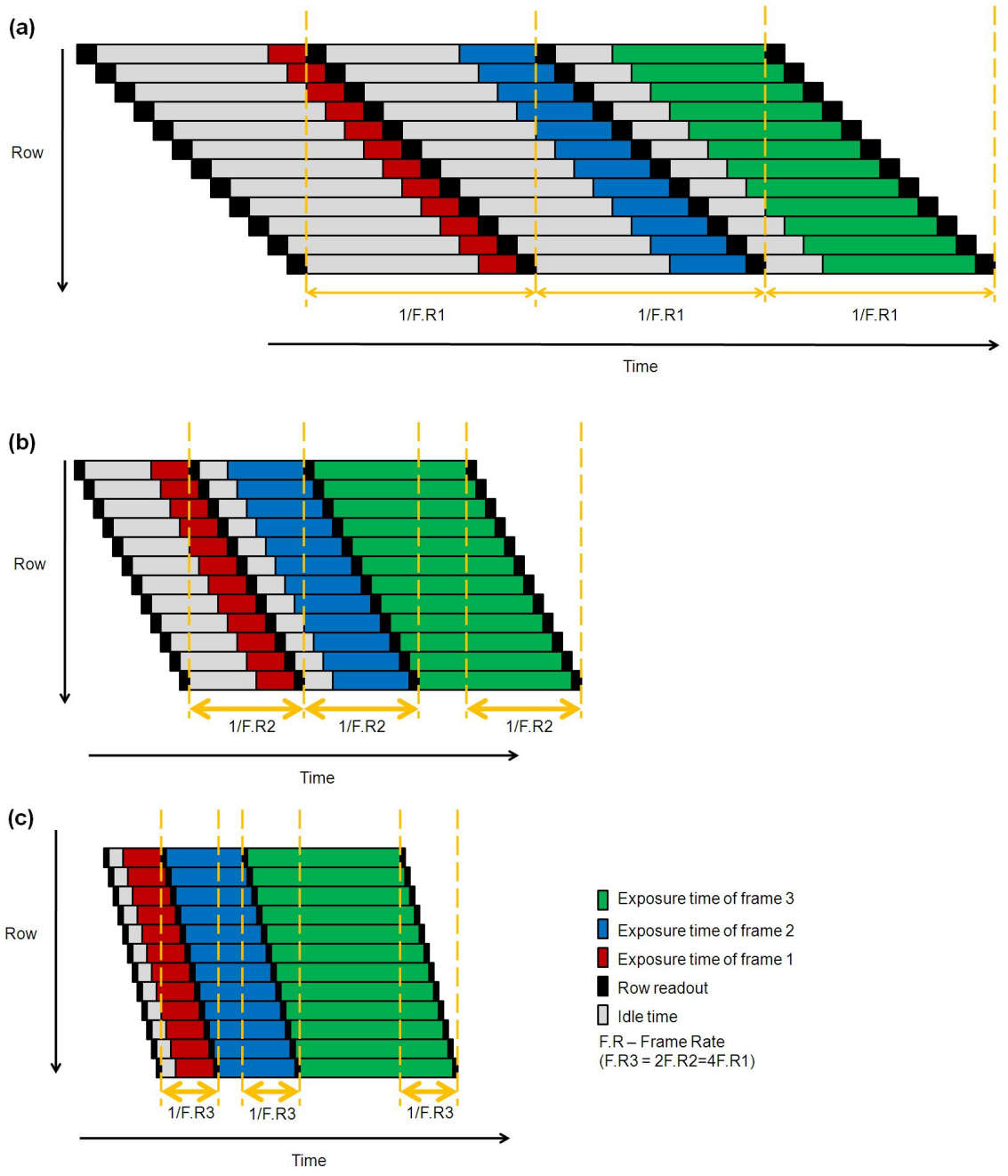


Figure 2.3: *Capturing three images with different exposures using conventional rolling shutter sensors with (a) low (b) mid and (c) high frame rates. If the exposure time is significantly shorter than the readout rate (top image), then the requirement that readout of frame N cannot begin until readout of frame N-1 has completed (this constraint is represented by the dashed vertical yellow lines) leads to a high percentage of idle time.*

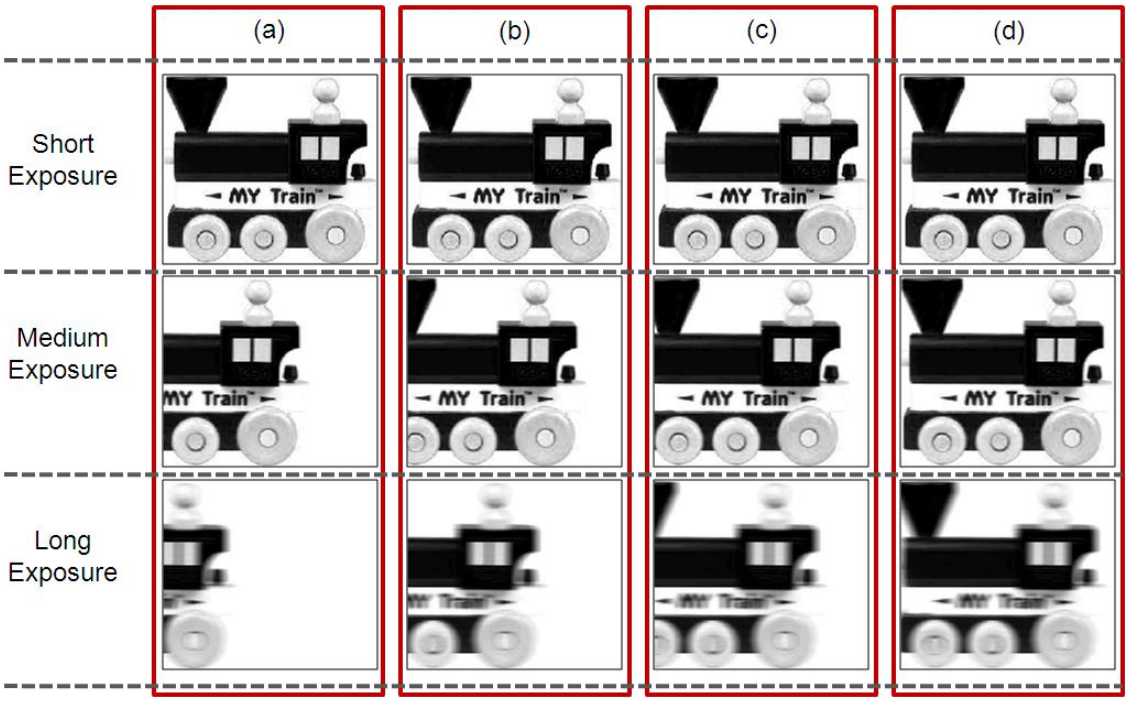


Figure 2.4: *Sample images of a moving object captured by sensors with (a) low (b) mid (c) high and (d) infinite frame rates. In each case, three images with different exposure times are taken sequentially. The rows of the figure represent different exposure times (top - short, middle -mid, and bottom - long). In this particular example, the frame rates of (a), (b), and (c) are 10fps, 20fps, and 40 fps, respectively. The exposure times for the short, medium, and long exposures are 1/960s, 1/240s, and 1/60s, respectively. The toy train is moving to the left at 960 pixels/s. The rolling shutter artifact is omitted here for simplicity. An interesting observation is that the three captured images are different from each other even in the case when a sensor with infinite frame rate is used.*

being used since the days of film-based photography, perhaps there is an alternative approach to capture multiple images with digital sensors.

2.2 Simultaneous Image Capture

Intuitively, the ideal way to capture multiple images would be capturing them simultaneously. In this scenario, any undesired changes in the scene would appear equally in all the captured images. For example, all captured frames would have the same handshake or object motion blur, and moving objects are in the same position in all frames. Therefore, the need to perform an image alignment or motion compensation is eliminated. One particular example of this imaging approach is Fuji's HDR super-CCD [Nakamura 06] in which each pixel contains two photodiodes, one more sensitive than the other, so that two images are captured simultaneously. The two captured images are correctly exposed to different regions in the scene and therefore they can be combined to form a HDR photograph. As the images are captured simultaneously, the reconstruction is insensitive to motion or varying ambient lighting conditions in the scene.

However, capturing multiple images simultaneously is not compatible with some computational photography techniques that use active illumination to modify the scene. For example, in flash/no-flash imaging [Petschnigg 04] [Eisemann 04], two exposures are made, one without a flash and the other one synchronized with a flash. These two captured images are combined to compute a final image that has shadowing and specular glow effects from the flash reduced so the mood of the scene is preserved

while in the meantime has dark regions of the scene correctly exposed. Obviously, the two images cannot be captured simultaneously in this case because the scene is being modified sequentially.

Apparently, there is a contradiction. On one hand, we want simultaneity to mitigate the differences between the multiple images due to undesired changes in the scene. On the other hand, simultaneity is not allowed because we want to take pictures of the scene that is being modified sequentially by us. In the next section, we describe an alternative imaging approach called time-multiplexed exposure that will be shown to resolve the above contradiction.

2.3 Time-Multiplexed Exposure

Time-multiplexed exposure is a generalization of conventional image capture. In this framework, an image is no longer constrained to be captured in a contiguous block of time. Instead, each exposure can be partitioned into pieces which can be interlaced with pieces of other exposures in any desired way. Figure 2.5 illustrates this concept, again using the example that three images with different exposure times (e.g. T1, T2, and T3) are to be captured. Under time-multiplexed exposure, frame 1, 2, and 3 now correspond to the sum of the sub-images captured in the red, blue, and green time slots, respectively.

The essence of time-multiplexed exposure is that it provides the flexibility to decide the centroid and the absolute time span of each image. Figure 2.6 illustrates this point by considering two capturing scenarios. The centroids of frame 1, 2, and 3

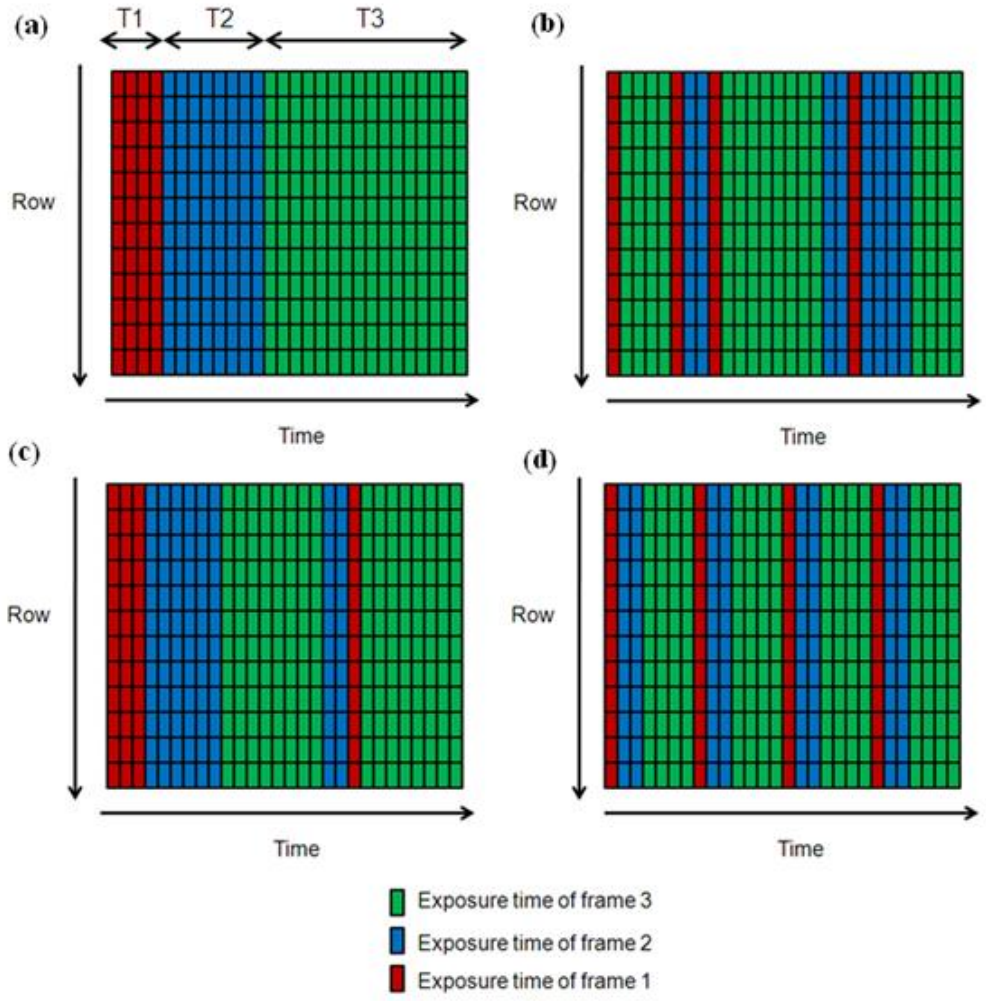


Figure 2.5: *Illustration of time-multiplexed exposure. (a)-(d) show four examples of time-multiplexed exposure. In each case, frame 1, 2, and 3 correspond to the sum of the sub-images captured in the red, blue, and green time slots, respectively.*

(shown as black dots in the figure) captured using the strategy given in Figure 2.6(a) are at T1/2, T1+T2/2, and T1+T2+T3/2, respectively. Therefore, the differences between centroids of frame 1 and frame 2 is (T1+T2)/2 and that between frame 2 and frame 3 is given by (T2+T3)/2. Now consider the capturing strategy shown in Figure 2.6(b). Assume each image is partitioned into N sub-images and let P_{i1}, P_{i2}, and P_{i3} be the centroids of the i^{th} sub-image of frame 1, 2, and 3 respectively. We have P_{i2}-P_{i1} = (T1+T2)/2N and P_{i3}-P_{i2} = (T2+T3)/2N. The difference between the centroids of frame 1 (C1) and frame 2 (C2) is given by:

C2 - C1 = (P_{12}+P_{22}+P_{32}++P_{N2})/N - (P_{11}+P_{12}+P_{31}++P_{N1})/N
= [(P_{12}-P_{11}) + (P_{22}-P_{12}) + (P_{32}-P_{31}) ++ (P_{N2}-P_{N1})]/N
= [N ((T1+T2)/2N)]/N
= (T1+T2)/2N

Similarly, the difference between centroids of frame 2 and frame 3 (C3) is given by:

C3 - C2 = (T2+T3)/2N

Therefore, the centroids of the three captured images are N times closer compared with those captured in Figure 2.6 (a) which corresponds to the case of conventional exposure using a sensor with infinite frame rate. Additionally, compared with the capture strategy in Figure 2.6(a), the difference in absolute time spans between frame

1 and 2 using strategy shown in Figure 2.6 (b) becomes (T2-T1)/N compared with T2-T1. Similarly, the difference in absolute time span between frame 2 and frame 3 becomes (T3-T2)/N compared with T3-T2.

The significance of the above results is that, by using the capture strategy in Figure 2.6 (b), we can in principle make the multiple frames arbitrarily similar to each other by increasing N, irrespective of the difference in relative exposures times of the frames. Therefore, time-multiplexed exposure not only can be used to minimize the difference due to offset of image capture in time (i.e. their centroids), but can also be used to minimize that due to different exposure times of the frames (i.e. their absolute time spans). Figure 2.7 shows sample images corresponding to the timing diagrams shown in Figure 2.6 (a), (b) and those that would be captured by conventional exposure using a sensor with a finite frame rate. Since the three exposures in Figure 2.7 (b) are now tightly interleaved and represent virtually the same span of time, undesired changes in the scene, such as motion in this example, are more evenly distributed between the three images. In particular, all frames captured using this strategy would have the same handshake or object motion blur, and moving objects would be in the same position in all frames. Therefore, this interleaving eliminates the need to align the frames or perform motion compensation after capture.

Furthermore, time-multiplexed exposure is compatible with multi-image computational photography that uses active illumination. For example, Figure 2.8 illustrates how flash/no-flash imaging [Petschnigg 04] [Eisemann 04] can be performed using time-multiplexed exposure. Again, by partitioning the flash and no-flash images into

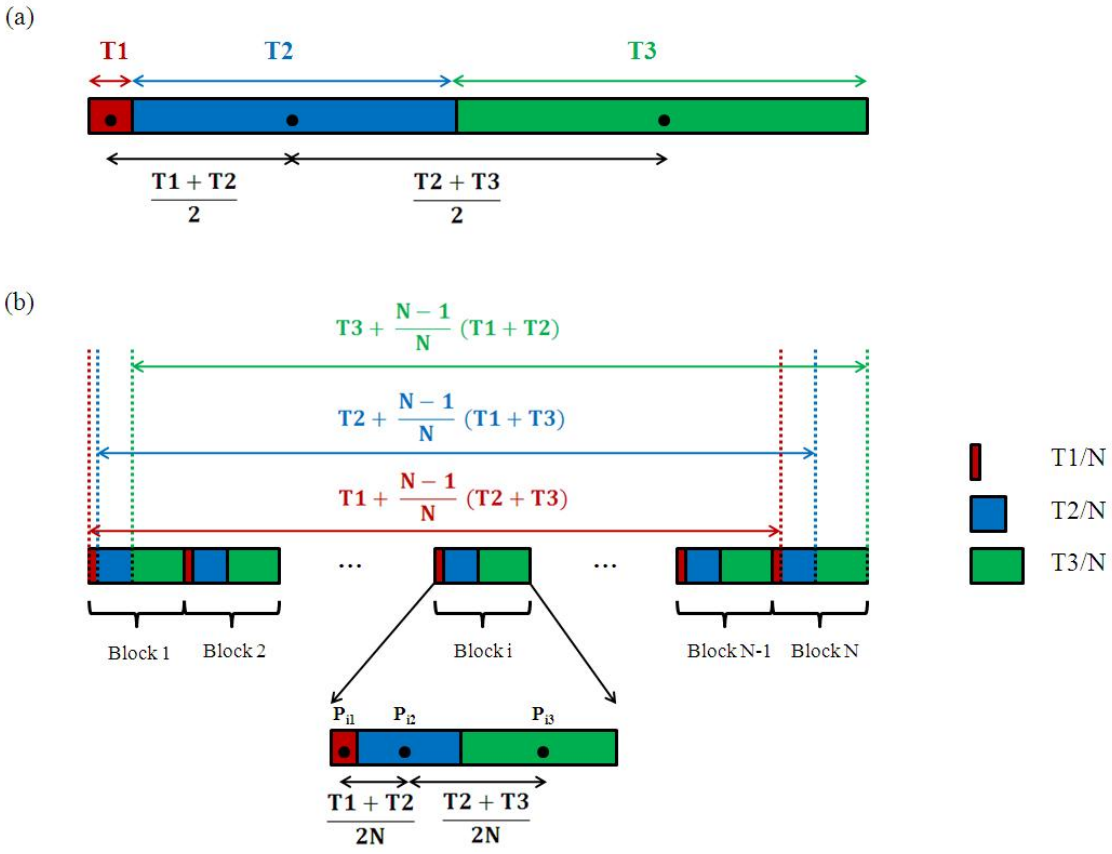


Figure 2.6: *Illustrating the flexibility of time-multiplexed exposure in modifying centroids and absolute time spans of multiple frames. (a) and (b) show two different capturing strategies under the framework of time-multiplexed exposure. In (a), three frames are captured back-to-back while in (b), each frame is partitioned into N equal pieces and the pieces of different frames are interleaved periodically into N blocks. The black dot inside a block indicates the location of the block's centroid. Compared with (a), the strategy in (b) reduces differences between centroids and absolute time spans of successive frames by a factor of N.*

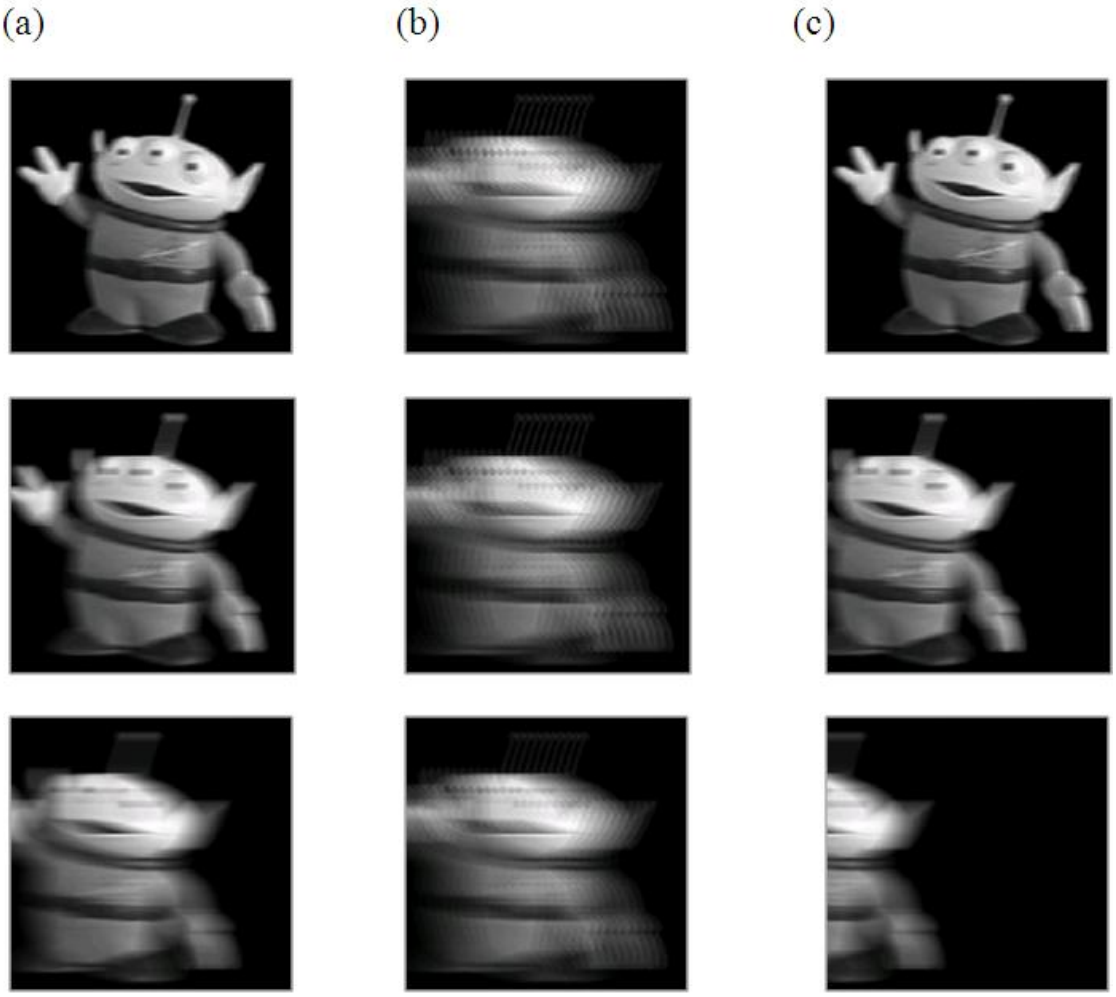


Figure 2.7: *Sample images using (a) capture strategy shown in Figure 2.6 (a), (b) capture strategy shown in Figure 2.6 (b), and (c) conventional exposure using a sensor with limited frame rate. The rows of the figure represent different exposure times (top - short, middle -mid, and bottom - long).*

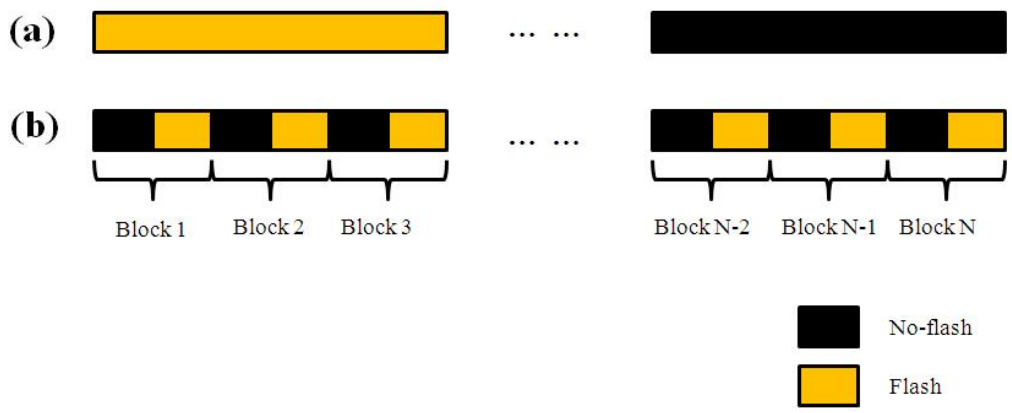


Figure 2.8: *Flash/no-flash imaging using (a) conventional exposure (b) time-multiplexed exposure. The centroids of the images in (b) are N times closer than that in (a) while the absolute time spans of the flash and no-flash images are the same. Effectively, the two images are captured virtually simultaneously in (b) despite the fact that the scene is being modified by the flash in one capturing one of the two images.*

N sub-images and interleaving them, centroids of the two images can be made arbitrarily close to each other by increasing N while the absolute time spans in this case are always equal to each other. The two images can therefore be made arbitrarily similar to each other except desirable differences caused by the flash. Similarly, more than one lighting condition can be interleaved in the same manner therefore allowing us to take multiple images under different lighting conditions virtually simultaneously.

2.4 Tradeoff in Time-Multiplexed Exposure

Using time-multiplexed exposure to minimize the differences between centroids and absolute time spans of images does come with a cost - it increases the absolute time span of each image and therefore makes the images more susceptible to motion blur. For example, the three images shown in Figure 2.7 (b) are similar to each other but

each of them also spans a longer absolute time compared with the case in Figure 2.7 (a) in which the centroids of the images are farther apart.

Whether the above tradeoff is justified depends on the application and the scene. For example, in video to be viewed by human eyes, motion blur is in fact desirable because it makes the video look much smoother and more natural whereas a sequence of images without motion blur would appear to be stroboscopic. In fact, modeling motion blur is important in animated sequences to remove temporal aliasing effects [Potmesil 83] but it remains one of the hardest challenges for current production renderers because of its mathematical complexity [Grünschloß 08]. Even for still photography, motion blur portrays the illusion of speed or movement among the objects in the scene and is often used by photographers as a creative element to produce artistic effects. Of course in situations when motion blur is not desirable, e.g. handshaking blur, we can simply use the capture strategy shown in Figure 2.6 (a) in which multiple images are captured back-to-back, an allowable situation under the framework of time-multiplexed exposure. In any case, having the flexibility to partition each exposure into pieces and interleave them in any arbitrary way allows us to choose the most suitable capture strategy for the application and scene at hand.

2.5 Example Use Cases

In this section, we compare conventional exposure with time-multiplexed exposure in four typical scenarios. To limit the scope of discussion, we assume that a normal sensor with limited frame rate (e.g. 30fps) is used in conventional exposure and that

four images are to be captured in each scenario. Also, we assume that a handheld camera (e.g. a smartphone) is used in each scenario so that handshaking is taken into account.

2.5.1 Low-light, low-dynamic range scene (e.g. restaurant with ambient light)

To avoid motion blur due to handshaking, each exposure should not exceed a maximum of 1/30s. Therefore, in conventional exposure using a normal sensor, each image is exposed for 1/30s and the total capture time is 4x1/30=2/15s. In this scenario, we cannot partition an image into two or more pieces in time-multiplexed exposure because doing so results in an image that spans more than 1/30s. Therefore, time-multiplexed exposure does not provide any advantage in this particular use case assuming technique such as flash/no-flash [Petschnigg 04] [Eisemann 04]is not being used.

2.5.2 Mid-light, low-dynamic range scene with fast motion (e.g. indoor sporting event)

In order to capture the fast motion, each capture should be no longer than 1/500s. In the conventional exposure using a normal sensor, since frame rate is limited to 30fps, it takes a total of 4x1/30=2/15s to capture the four images. On the other hand, when time-multiplexed exposure is used, we can still expose each image for 1/500s but they can be captured back-to-back and the total exposure time is then reduced

to 4x1/500 = 1/125s. The 4 images can be aligned and averaged to compute a final image which has signal-to-noise ratio improved. Since time-multiplexed exposure, by behaving like a conventional sensor with an infinite frame rate in this case, shortens the total capture time by 16X, the subsequent image alignment is made much easier.

2.5.3 Low-light, high-dynamic range scene (e.g. dark room with sunfilled window)

In this case, typical exposure times of the four captures are 1/60s, 1/125s, 1/250s, and 1/500s. Again, due to the limited frame rate at 30fps, the total capture time is 4x1/30 = 2/15s. In the case of time-multiplexed exposure, the total exposure time is shortened to 1/60+1/125+1/250+1/500 $\approx$ 1/30s. In this indoor scene, most people would expect 1/60s to be the maximum exposure time for a single image. Therefore, time-multiplexed exposure offers an advantage here because we can keep the first exposure to be 1/60s, then partition the following three exposures and interlace them while making sure that the time span of each is shorter than 1/60s. In this case, we do not increase motion blur of any capture while minimizing the distances between centroids of the last three captures as well as the differences of their total time spans. Compared with conventional exposure, time-multiplexed exposure reduces the total capture time by 4X in this scenario making the image alignment, if needed, much easier.

2.5.4 Bright, high-dynamic range scene (e.g. outdoor with objects in shadow)

Typical exposure times in this case are 1/2000s followed by 1/1000s, 1/500s, and 1/250s. In conventional exposure using a normal sensor, it takes a total of 4x1/30 = 2/15s for the entire capture. In this scene, relative to the standard outdoor exposure of 1/2000s that most people would expect, we can choose not to partition any exposure in time-multiplexed exposure but simply use it to capture the four frames back-to-back. This shortens the total exposure time to 1/2000+1/1000+1/500+1/250 = 3/400s, a 18X improvement in terms of image alignment difficulty compared with the case of conventional exposure. Alternatively, since each exposure is relatively short, if motion blur of fast-moving objects can be tolerated, some or all of the exposures can be partitioned and interlaced to bring their centroids closer together and total time spans similar to each other and thus eliminating the need to do any image alignment altogether.

2.6 Implementation

Memory is an indispensable element in time-multiplexed exposure because sub-images need to be stored and accumulated during an exposure. In principle, there are two options in placing the memory - either inside or outside the pixel array. In this section, we will examine both architectures in details. We will first see that the architecture that uses external memories has prohibitive requirements on frame rate, data rate,

and power consumption and therefore it is an infeasible option if current technology is used. Next, we will describe our chosen architecture that uses local or pixel-level memories and explain the reason why it is a preferable option in implementing time-multiplexed exposure.

2.6.1 Fast Sensor and External Memory

Figure 2.9 shows an architecture in which memory is located outside the pixel array. In time-multiplexed exposure, each of the images to be captured can be partitioned into smaller sub-images. Since we do not want time gaps to exist between the sub-images, this architecture limits the shortest exposure of a sub-image by the frame rate of the sensor. For example, a 2000fps sensor is required if a 1/2000s exposure time is desired. Although sensors with such high speeds have been reported in the literature [Krymski 03a] [Furuta 06], their low spatial resolutions make them unsuitable for photography applications.

Currently, the architecture of choice for a high speed CMOS imager is the so-called column-parallel architecture [Nakamura 06] [Ohta 07] [Kawahito 10] shown in Figure 2.10. In this architecture, an image is readout one row at a time. During one row time, pixels of the same row are read out to the column circuitries simultaneously and processed in parallel. Each row processing includes programmable signal amplification, sample-and-hold operation, and finally analog-to-digital conversion. The processed signals are then stored in column line memories before they are read out through one or multiple output ports to the external memory.

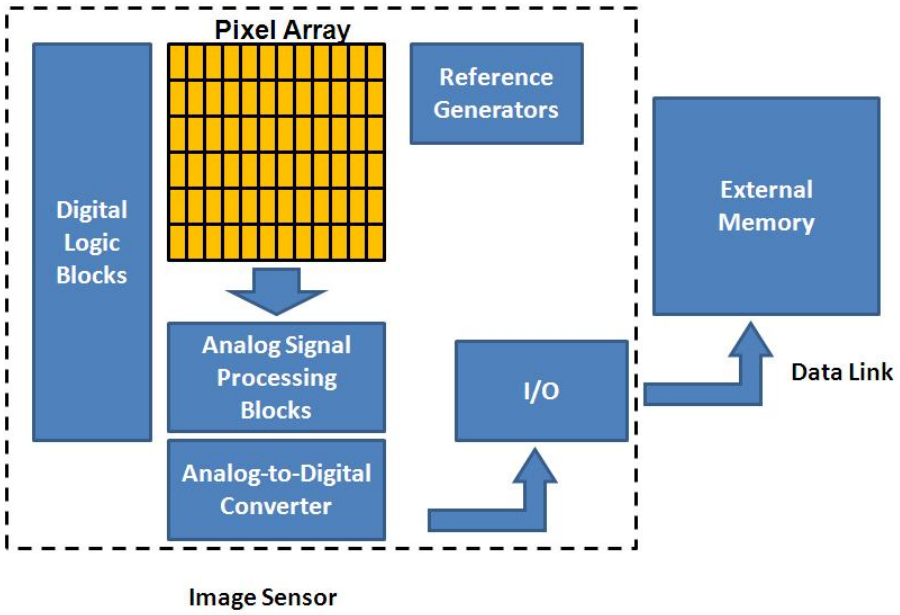


Figure 2.9: *Implementing time-multiplexed exposure using external memories.*

As we can see, increasing the sensor's frame rate therefore requires reducing the row readout time. Figure 2.11 shows the relationship between row readout time and the number of pixels at different frame rates for image sensors with two standard aspect ratios. Currently, a consumer grade image sensor with 8 megapixels (Mpix) runs at a maximum of 30fps while the row time of a conventional 2Mpix, 60fps HDTV image sensor is about 15μs. As we can see from the figure, a 2000fps sensor with 8Mpix in particular would require more than 60X reduction in row time compared with the commercial sensors. Referring to Figure 2.10(b), it requires speeding up the column circuitries. However, even if we assume the circuits can run arbitrarily fast, the row time is fundamentally limited by the time required for signal being readout from the pixel to settle at the column line which always has finite resistance and capacitance

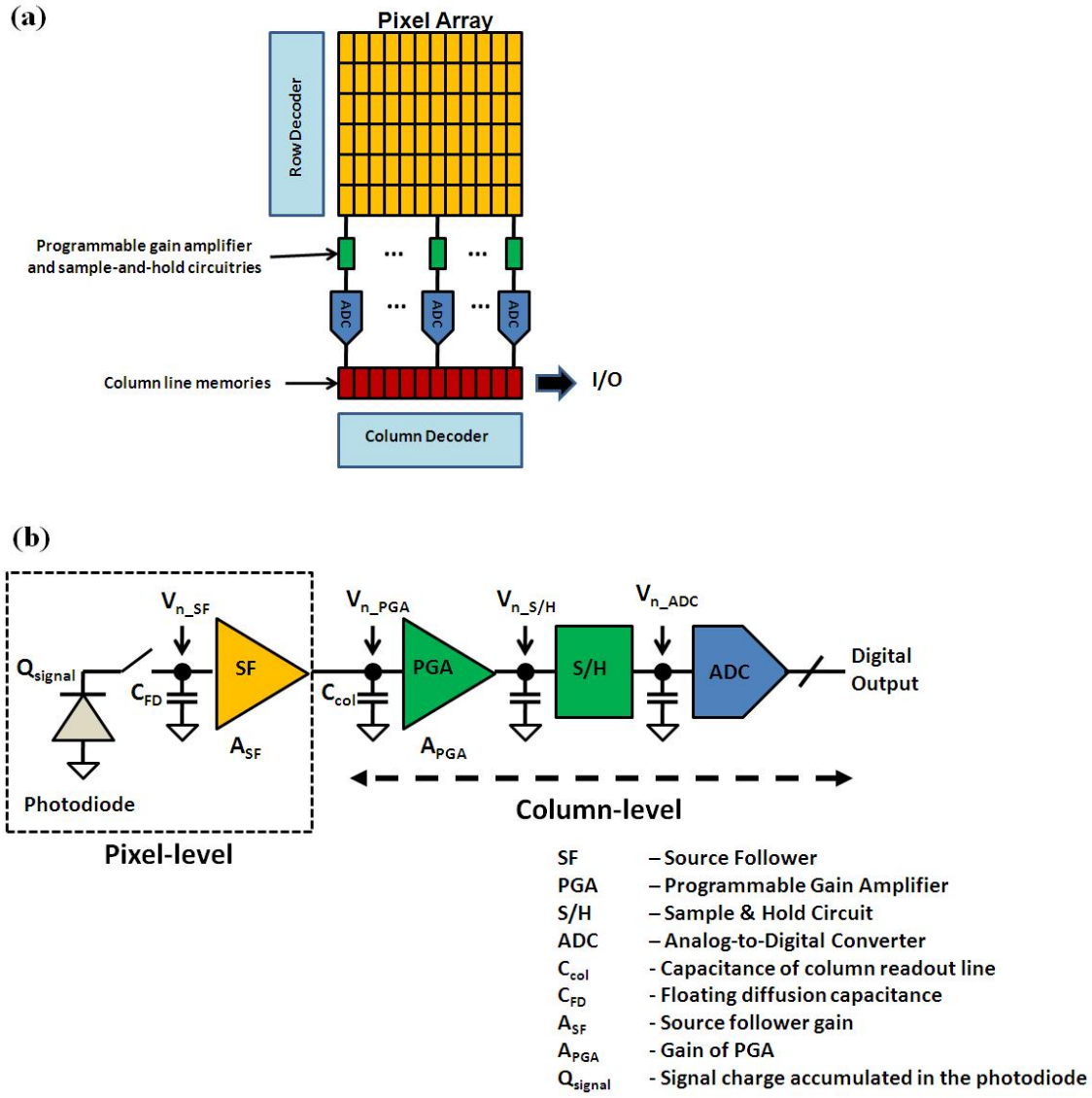


Figure 2.10: *The column parallel CMOS imager architecture. (a) Whole array (b) One column*

(C_{col}) and therefore a corresponding RC delay.

In fact, even if we have a sensor with infinite frame rate, this architecture poses a high bandwidth requirement on the data link to the external memory. Assume each of the M images to be captured are partitioned into N sub-images and the pixel array consists of K pixels. Totally MxNxK pixels need to be readout from the pixel array to the external memory within an exposure time T. Figure 2.12 shows the relationship between data rate and the number of pixel with respect to the number of sub-images N. For the situation described in Section 2.5.4, we want to capture 4 images (M=4) with no partition (N=1) and assume the pixel array has 8Mpix (K=8x10^6), each pixel has 10bits, and we need to readout all these data from the pixel array within T=1/125s. Then, the data rate would be around 48Gbps - a number which has already exceeded the state-of-the-art prototype sensors [Bogaerts 11] [Toyama 11]. Instead, if we want to partition each image into 8 sub-images (N=8), the number becomes impractically high at 384Gbps.

Another important limitation of this architecture is the total power consumption which increases with pixel count, frame rate, and data rate. As shown in Figure 2.9, the image sensor consists of various circuit blocks. The digital blocks consist of timing generators, column and row decoders, and other circuits to control the operation of the image sensor. The analog signal processing blocks consist of programmable gain amplifiers, sample-and-hold circuitries. To realize a fully digital interface, an analog-to-digital converter is incorporated in the same image sensor chip. Other circuits include for example reference generators. While the absolute power consumption

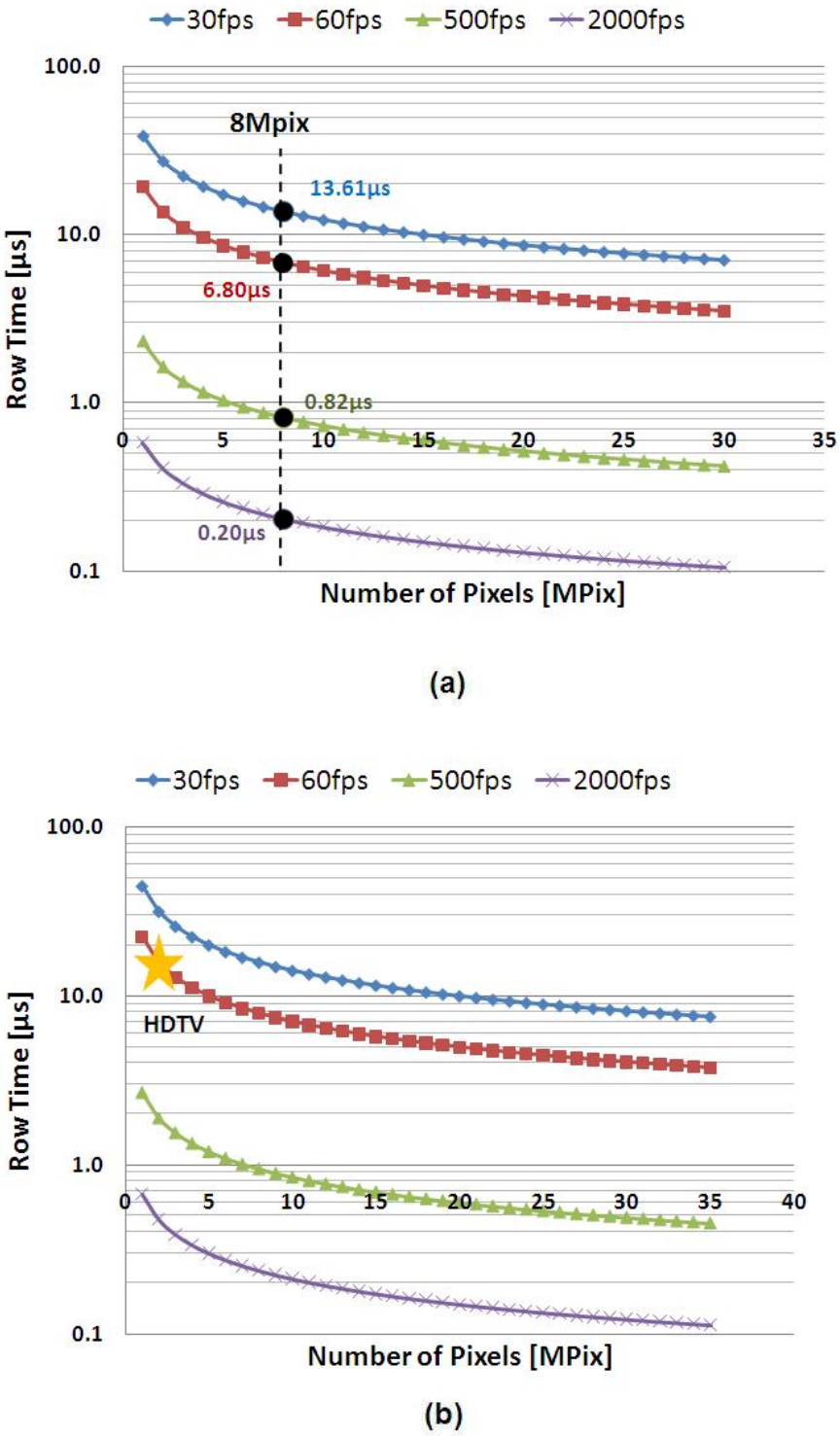


Figure 2.11: *Relationship between the row readout time and the number of pixel parameterized by frame rates. (a) 4:3 aspect ratio (b) 16:9 aspect ratio.*

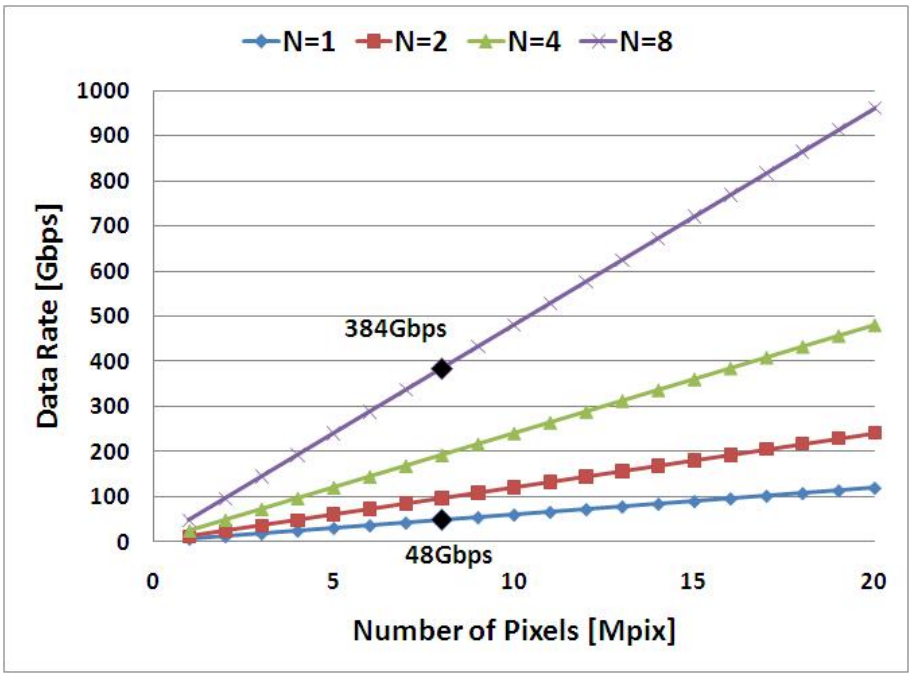


Figure 2.12: *Relationship between data rate and the number of pixel parameterized by the number of sub-images (Each pixel has 10bits).*

of each block depends heavily on the chosen circuit topology, it is universally true that the faster the circuits or the higher the data rate, the more the power needs to dissipate at a given process technology node. For example, the two state-of-the-art sensors that deliver 36Gbps [Bogaerts 11] [Toyama 11] consume around 3W while a typical image sensor dissipates below 500mW. To estimate the total power that a hypothetical sensor would consume when performing time-multiplexed exposure, we use the data in [Toyama 11]. The power consumption excluding I/O is around 1.32mW/[(Mpix)(fps)] while that of the I/O is around 5.75mW/Gbps. Figure 2.13 shows estimated power consumption with I/O excluded with respect to number of pixels and frame rate. Consider again the example in Section 2.5.4, a 8Mpix sensor with 2000fps would consume 21W. If we further assume that the data rate is around 48Gbps, the total power consumption would become 21.28W - an impractically high number. Such large power consumption is detrimental to battery life. Also, without an efficient thermal design of sensor package, the large amount of heat dissipated increases dark currents being generated in the image sensor array [Theuwissen 95] and therefore worsens the image quality. Note also that in reality, we need to take into account powers consumed in the external memory which makes the total power consumption of the system even worse.

Various amounts of noise are added to the signal when the later goes along the signal readout chain as shown in Figure 2.10(b). In the figure, V_{n_sf}, V_{n_PGA}, $V_{n\_S/H}$, and V_{n_ADC} represent the input-referred noises added by the source follower, PGA, S/H, and ADC, respectively. When all these noises are referred back to the input

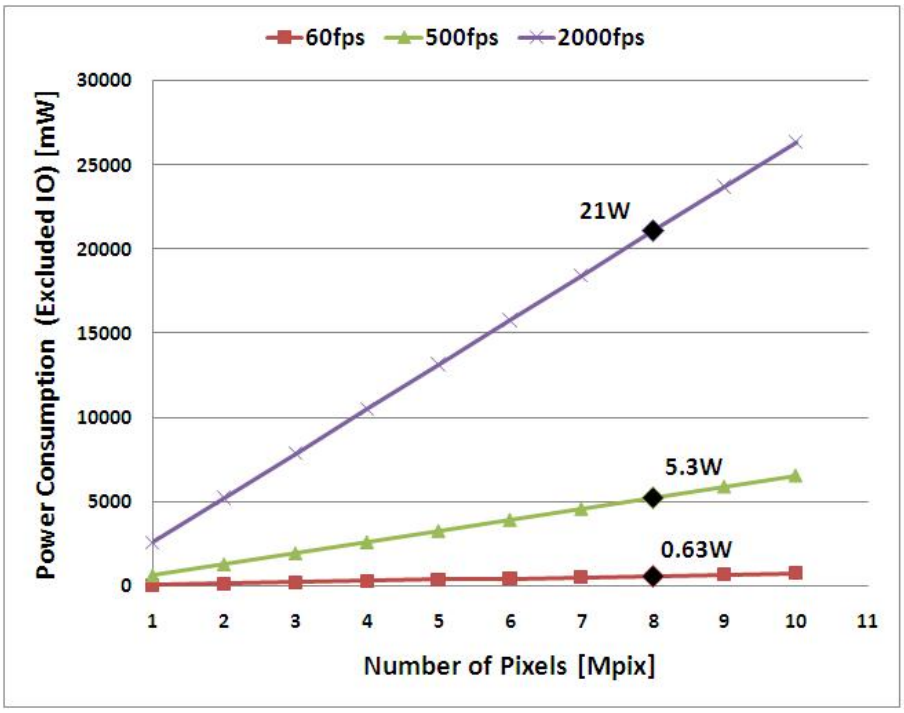


Figure 2.13: *Power consumption (excluding IO) with respect to the number of pixels and frame rates.*

of the source follower, also known as the floating diffusion (FD) where signal charge (Q_{signal}) is converted to voltage, the input-referred noise or read noise in terms of electrons is given by:

$$R^2 = \frac{1}{(\frac{q}{C_{FD}})^2}\left[V_{n_{SF}}^2 + \left(\frac{V_{n\_PGA}^2}{A_{SF}^2}\right) + \left(\frac{V_{n\_S/H}^2}{A_{SF}^2 A_{PGA}^2}\right) + (\frac{V_{n\_ADC}^2}{A_{SF}^2 A_{PGA}^2})\right]$$

where q = 1.6x10^{-19}C is the elementary charge.

As we can see from this formula, the larger the (q/C_{FD}), also known as conversion gain of the pixel measured in μV/electrons, the smaller the read noise. The performance of the architecture described in this section is fundamentally limited by this read noise because it is added to each sub-image every time it is readout from the pixel array. Consider the example that we partition an image equally into N sub-image and read them out separately. In this case, the signal of each sub-image is N times smaller and therefore the signal-to-noise (SNR) ratio of each sub-image is given by:

$$SNR = \frac{\frac{Q_{signal}}{N}}{\sqrt{R^2 + (\frac{Q_{signal}}{N})}}$$

where the second term in the denominator represents the photon shot noise which is proportional to the square root of the signal [Janesick 07]. As shown in Figure 2.14, the more we partition an image (i.e. the larger the N), the smaller the SNR

of each sub-image. Summing or averaging the sub-images after they have been read-out improves the SNR by $\sqrt{N}$ times [Janesick 07] as shown by the dashed lines in the figure. Nevertheless, because of the finite read noise, the SNR of the image is fundamentally smaller than if the image has not been partitioned (i.e. N=1). This limitation, however, is not as severe as the issues due to frame rate, data rate, and power consumption because of the continual improvement in read noise. Current image sensors can typically achieve 1-2e^- read noise [Krymski 03b] and sub-electron read noise has recently been reported [Lotto 11].

2.6.2 Pixel-Level Memory

Given the issues of the architecture described in the previous section, we explored another option to implement time-multiplexed exposure through the use of local or pixel-level memories. Local memories have previously been suggested to benefit image formation to the extent that there has been a hypothesis that the peripheral visual pathways are served by local memory circuitry [Wandell 02]. In this architecture, the N intermediate sub-images are added and stored inside pixel-level analog memories. Since the memories are at close proximity to the photodiodes and no signal processing or data conversion is involved, sub-images can be transferred to the memories very rapidly. Therefore, this architecture does not require a high frame rate sensor as in the previous architecture. Another advantage of this architecture is that, because sub-images are summed in the memories before readout, the amount of information transferred out from the pixel array is N times smaller than the previous architecture

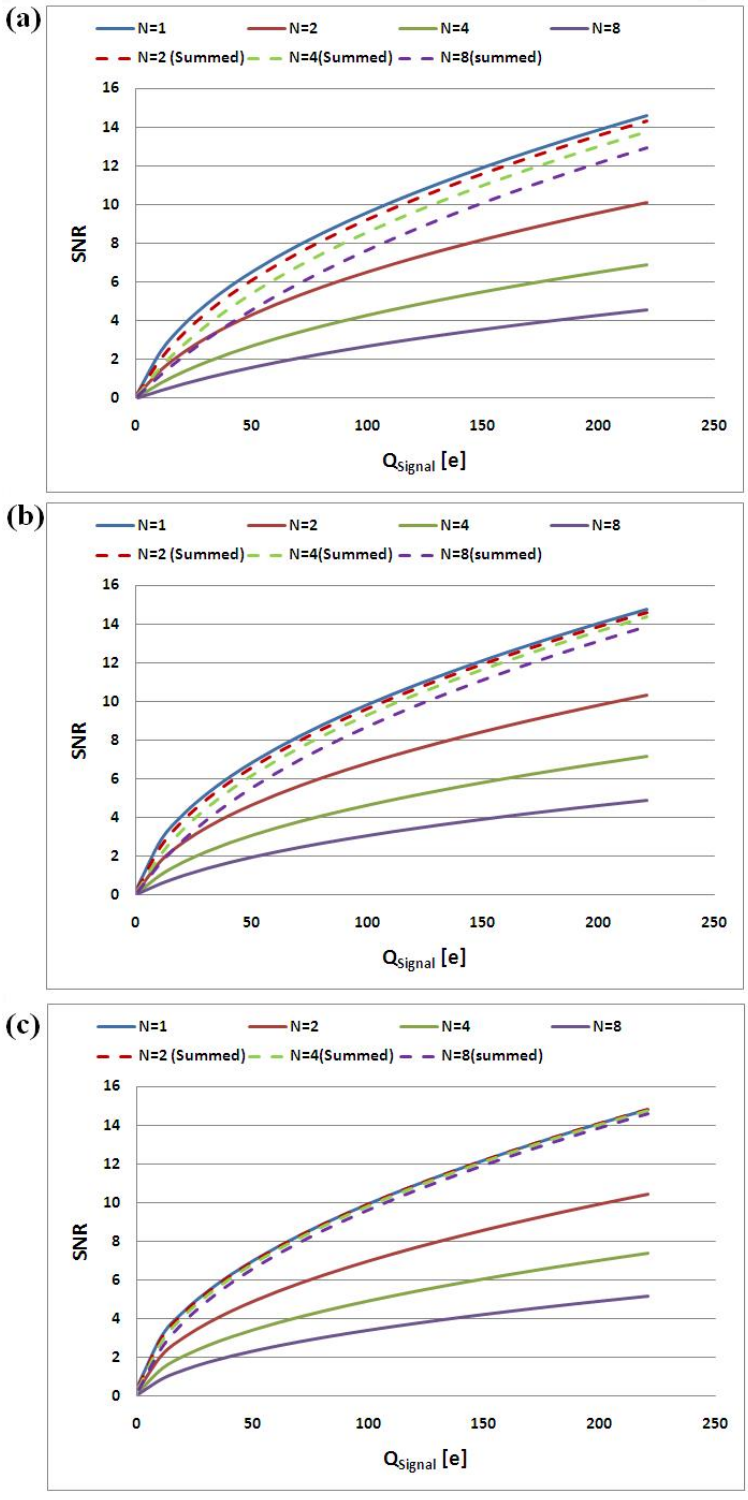


Figure 2.14: *Signal-to-noise ratios (SNR) versus signal levels. (a) $R = 3e^-$ (b) $R = 2e^-$ (c) $R = 1e^-$.*

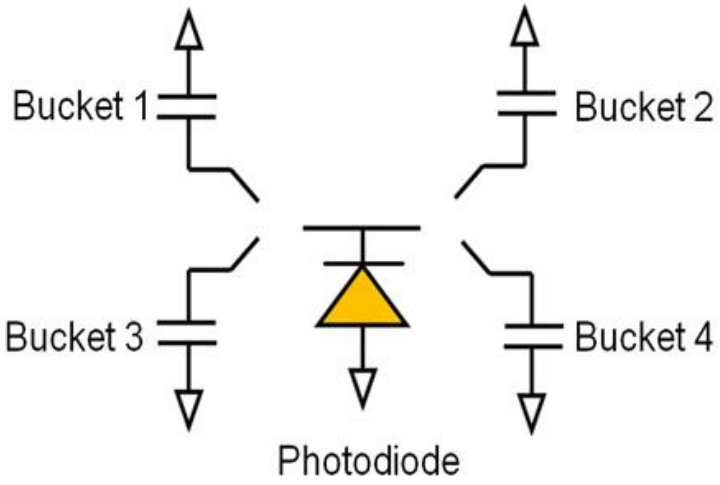


Figure 2.15: *A conceptual view of a multi-bucket pixel.*

and hence relatively low bandwidth data are output. As a result, the power consumption of this architecture is much lower and the engineering challenges of building the supporting system are much easier. In fact, it looks similar to today's systems. Furthermore, since sub-images are added within the local memories before readout, which can be a noiseless process, this architecture does not suffer from accumulating read noise like the case in the previous architecture.

In this architecture, each pixel needs to have two or more analog memories in order to perform time-multiplexed exposure. Figure 2.15 shows a conceptual view of a pixel with multiple analog memories. We will call this architecture a multi-bucket pixel. It consists of 1) a photodiode to convert light into charge 2) multiple buckets to accumulate photo-generated charges and 3) switches that are programmable by the user such that we can control which light goes into which bucket. Whereas frames

are acquired sequentially in a conventional sensor, in a multi-bucket sensor photo-generated charges in a photodiode can be transferred and accumulated in the buckets in any chosen time sequence during an exposure so multiple frames can be acquired in a time-multiplexed manner.

The advantages of this architecture, however, come with tradeoffs. After finishing one time-multiplexed exposure, all the buckets need to be readout before another round of time-multiplexed exposure can start. As a result, as shown in Figure 2.16, time gaps exist between the captures of multiple frames with the lengths of the gaps being inversely proportional to the frame rate of the sensor. These time gaps are important for videography but not still photography. Another challenge of this architecture is that a multi-bucket pixel without careful engineering would be bigger than a regular pixel because the buckets need to occupy extra pixel areas. As discussed in Section 1.1, small pixel size is critical for smartphones because it allows sensors with small form factors to be used. In the next chapter, we will discuss how we designed a multi-bucket pixel efficiently in current CMOS technology.

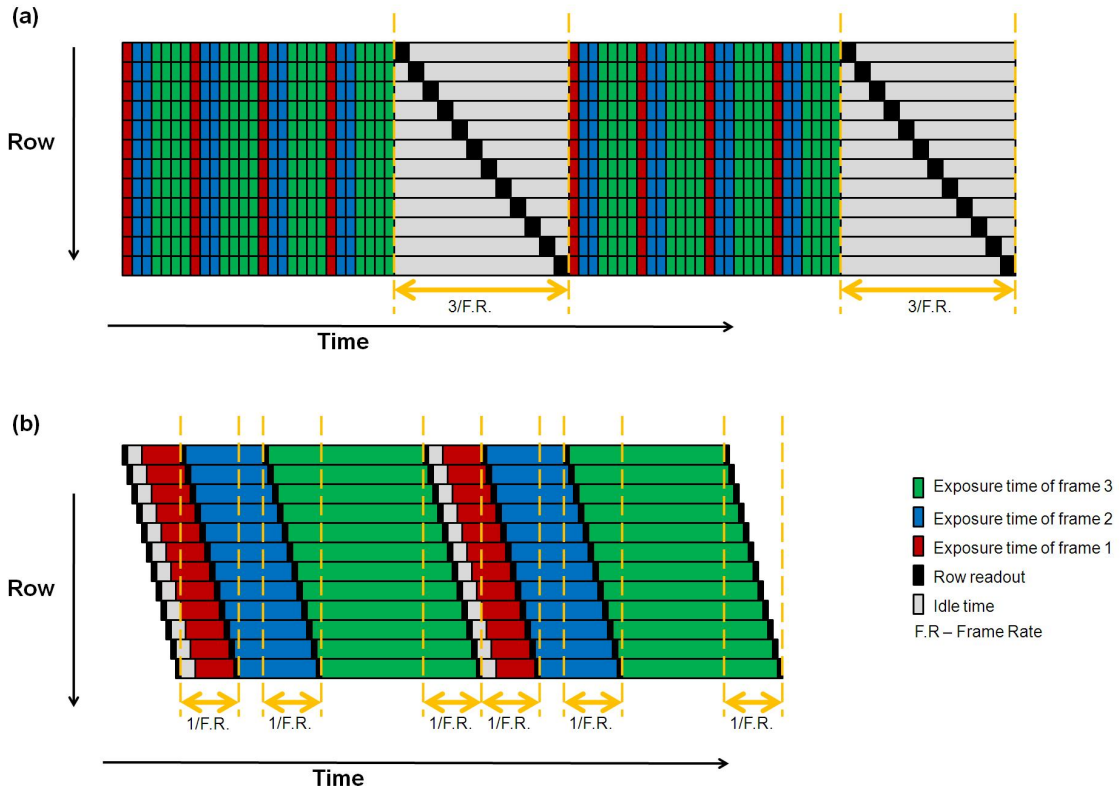


Figure 2.16: *Capturing two sets of multiple frames using (a) time-multiplexed exposure and (b) sequential exposure using a conventional rolling shutter sensor. In (a), since the buckets need to be readout before the next round of time-multiplexed exposure can start, time gaps exist between the two sets of multiple frames.*

Chapter 3

Multi-Bucket CMOS Pixel Design

In the previous chapter, we have seen that an efficient way to implement time-multiplexed exposure is through the use of a multi-bucket pixel that has multiple analog memories. Pixels in conventional image sensors, however, are either memory-less (rolling shutter pixel) or have only one memory (global shutter pixel) [Nakamura 06] and therefore they do not support time-multiplexed exposure.

Nevertheless, sensors with multi-bucket pixels do exist in the literature for special applications. For example, pixels with two memories, commonly known as lock-in or demodulation pixels, have been used to detect amplitude modulated light [Yamamoto 06], including time-of-flight 3D imaging [Kawahito 07] [Kim 10] [Stoppa 10], HDR imaging, motion detection [Yasutomi 10], etc. Typical implementations of a pixel memory include a floating diffusion [Yamamoto 06] [Kawahito 07] [Kim 10] [Stoppa 10], a MOS capacitor [Meynants 11], and most recently a pinned photodiode [Yasutomi 10]. However, pixels reported to date either have large sizes, low fill factors,

high kTC noise, or cannot be easily scaled to include multiple memories. Therefore, they are not suitable for photography applications.

In this chapter, we first review the basics of the CMOS pixel technology. Then, we discuss desirable features a multi-bucket should have followed by a review of some bucket implementations in the literature. Next, we analyze a bucket from a functional perspective. Finally, we present a new device structure which forms the basis of our multi-bucket pixels.

3.1 Basics of the CMOS Pixel Technology

In this section, we review basics of the CMOS pixel technology to provide a basis to discuss design challenges of a multi-bucket pixel. We focus on the four-transistor (4T) CMOS imager pixel since it is the dominant CMOS pixel technology used in almost all of the cameras and cell phones that are being shipped today.

Figure 3.1(a) shows the cross section and top view of the pixel. Light falling on the pixel opening is converged by the μlens sitting on the top of the pixel. The light then passes through the color filter array (CFA) and the dielectric stack inside which metals (e.g. M1 and M2) are routed. The pinned photodiode (PPD) inside the epitaxial layer (P-epi) collects and converts the light into charges which are accumulated inside the PPD before readout. The doping concentration of the P-epi is very low and therefore the depletion layer of the PPD can reach the p-type substrate (P-sub) edge to increase the photo-conversion volume and hence charge collection efficiency [Nakamura 06]. The PPD collects charges whereas output signal of the pixel is in

voltage. The charge-to-voltage conversion takes place at the floating diffusion (FD) node. The capacitance of the FD (C_{FD}) is related to the pixel's conversion gain (CG) by:

$$CG = \frac{q}{C_{FD}} \qquad [\text{Unit: } \mu V/e]$$

which essentially measures how much μV change is obtained by one electron at the FD. As we have described in Section 2.6.1, the smaller the C_{FD}, or the higher the CG, the smaller the read noise. Besides the PPD and the FD, the pixel has four transistors including a transfer gate (TX), a reset transistor (RST), a source follower (SF), and a row select transistor (ROW). The drains of RST and SF transistors are both connected to the pixel power supply (Vpix). Signal at the FD is readout from the pixel to the column readout chain through a column line (Col) connected to the source of ROW transistor.

As we can see in Figure 3.1(b), the transistors and the FD occupy some of the pixel areas. In fact, they can be considered as overhead because they reduce the pixel's fill factor which is defined as the ratio of the PPD area and the pixel area. Such a reduction of PPD area has two negative effects. First, it reduces the maximum amount of electrons that can be accumulated in the PPD, also known as the PPD's full-well capacity. Second, it lowers the light sensitivity of the pixel. To remedy this, the μlens at the top of the pixel has to be optimized to condense light onto the PPD as much as possible.

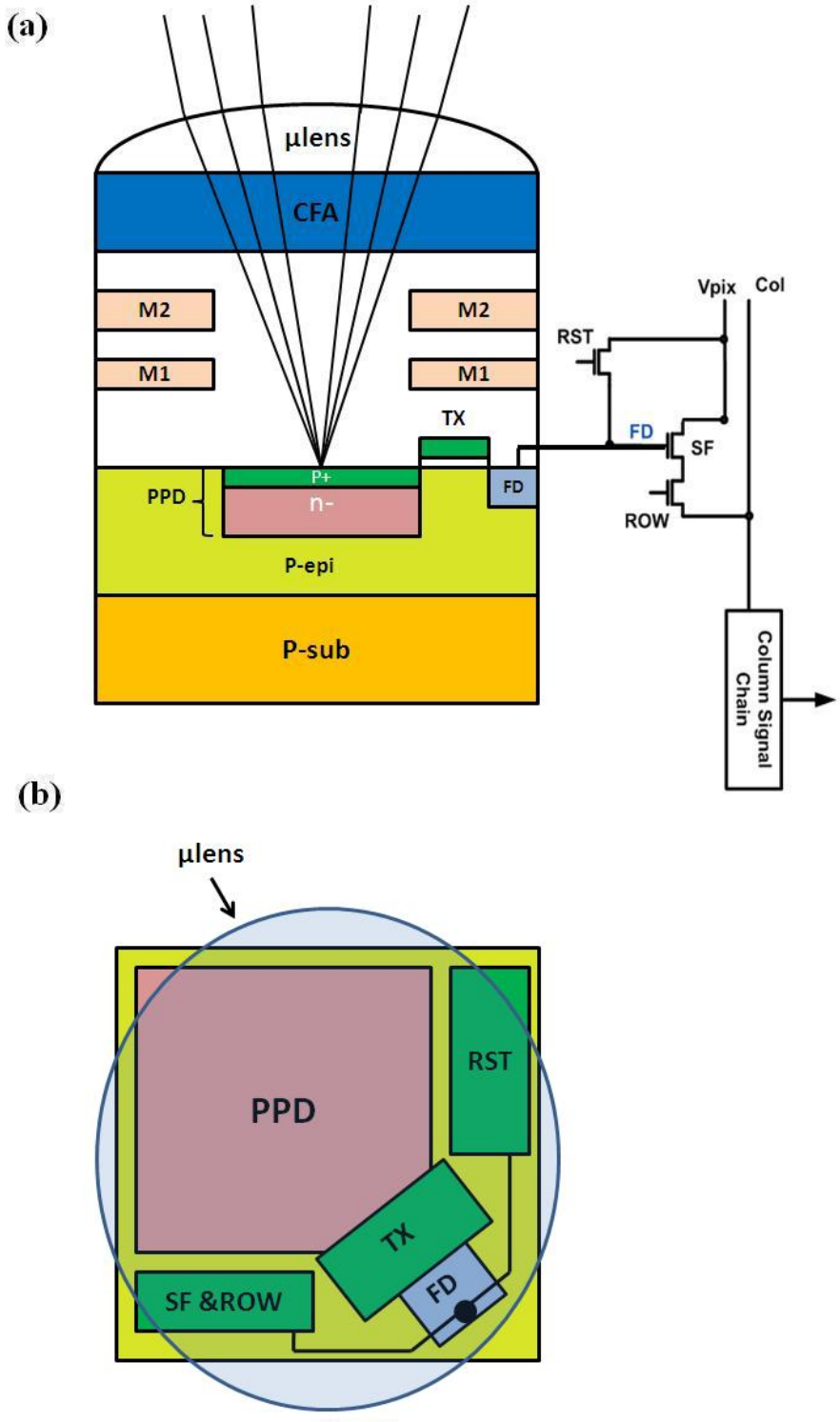


Figure 3.1: *The 4T CMOS imager pixel. (a) Cross-sectional view (b) Top view.*

3.2 Desirable Features

As we have seen in the previous section, the four transistor (4T) pixel is memoryless because the light sensing and charge accumulation regions are located at the same place - the pinned photodiode. Although a multi-bucket pixel is significantly different from such a memoryless pixel, there are features of the 4T pixel that we would like to maintain when designing our multi-bucket pixels. In this section, we discuss those desirable features in addition to some features which are unique to a multi-bucket pixel.

3.2.1 Correlated Double Sampling

One of the reasons that makes the 4T pixel so popular is because it supports true correlated double sampling (CDS) to cancel kTC reset noise. Figure 3.2 and 3.3 show a circuit schematic of the 4T pixel and its readout timing diagram, respectively, to illustrate the CDS operation. At the beginning of a readout, the floating diffusion (FD) is first reset to a high potential by turning on the reset transistor (RST). After the FD is reset, the RST transistor is turned off. Similar to any switched-capacitor circuit, at that moment thermal noise of the RST transistor appears at the FD node [Razavi 01]. Figure 3.4 shows an equivalent circuit of the reset operation. When the RST transistor is on, the power spectral density of the thermal noise is given by:

$$S_v(f) = 4kTR \quad [\text{Unit: } V^2/\text{Hz}]$$

where k is the Boltzmann's constant, T is the absolute temperature, and R is the

resistance of the RST transistor when it is on. When the RST transistor is turned off, this noise is sampled and held at the FD. The resulting noise power can be calculated by integrating the thermal noise power over all frequencies [Razavi 01] as given by:

$$V_{noise}^2 = \int_0^\infty \frac{4kTR}{1 + (2\pi fRC_{FD})^2} df = \frac{kT}{C_{FD}}$$

The noise charge is therefore given by:

$$Q_{noise}^2 = C_{FD}^2 V_{noise}^2 = kTC_{FD}$$

Since Q_{noise} is a random variable that varies from pixel to pixel, this reset noise or so-called kTC noise, contributes to the sensor's noise floor. In CDS, this reset level is sampled to an external capacitor through asserting the signal SHR. Next, TX is turned on to completely transfer the charges accumulated in the PPD to the FD. This signal charge (Q_{signal}) causes an additional drop in the FD potential. This signal level is sampled to another external capacitor by asserting the signal SHS. The true signal that represents the accumulated charge in the PPD can therefore be obtained by simply subtracting the two sampled signals regardless of the fluctuating reset levels across the imager array. Given the effectiveness of CDS in reducing the noise floor, one desirable feature a multi-bucket pixel should have is to preserve this CDS capability.

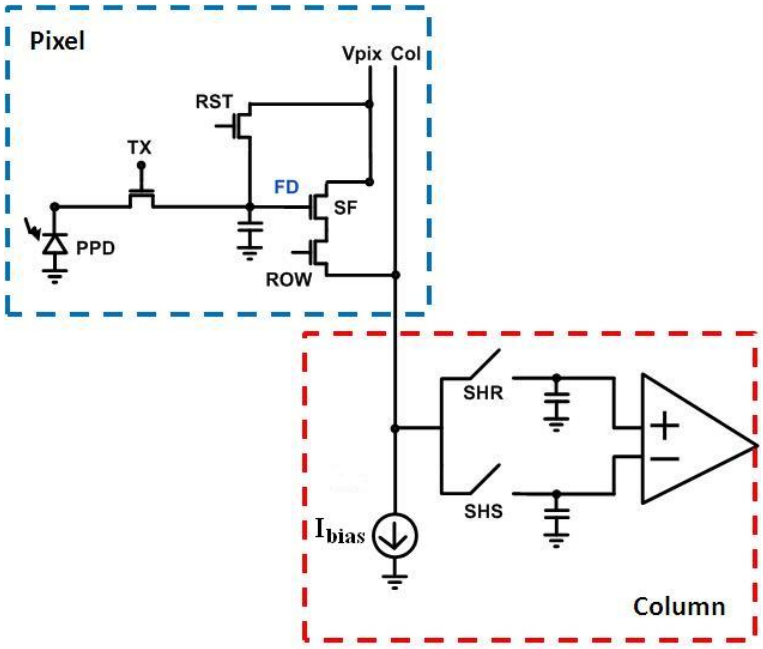


Figure 3.2: *Circuit schematic of a 4T CMOS pixel.*

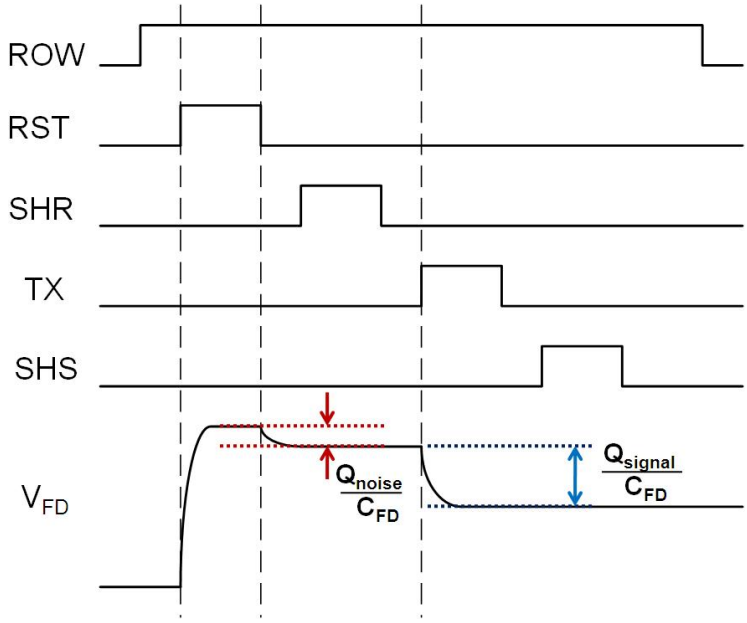


Figure 3.3: *4T pixel readout timing diagram to illustrate CDS. Q_{noise} - Noise charge injected into FD. Q_{signal} - Signal charge transferred from the PPD to the FD.*

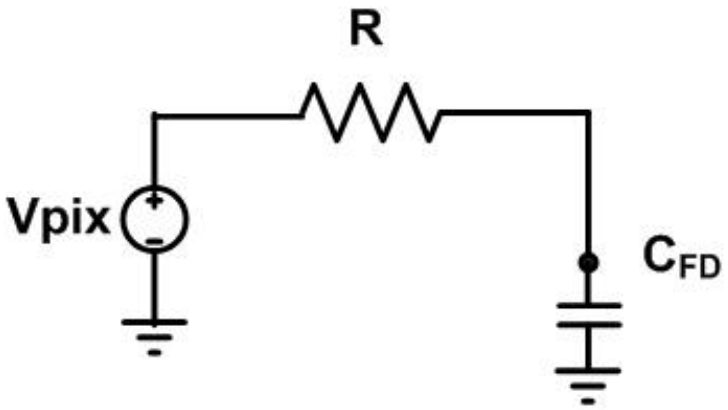


Figure 3.4: *Equivalent circuit of the reset operation.*

3.2.2 Low Dark Current Sensing and Storage

Another reason that the 4T pixel is dominant in recent CMOS image sensors is because the PPD has an extremely low dark current, for example, as low as 50pA/cm^2 at 60^{o}C [Nakamura 06]. Dark current is any undesirable leakage current that is integrated as charge at a charge storage node inside a pixel when no light is present. There are various sources of dark current but the one that contributes the most in a pixel is surface-generated dark current [Theuwissen 95]. First introduced to the interline transfer CCD image sensor [Teranishi 84], the PPD has its surface covered by a p$^+$ layer biased at ground. This layer allows holes to fill up the interface states at the surface which significantly reduces surface-generated dark current [Theuwissen 06]. Therefore, it is desirable to use a PPD as the photo-sensing element or even the bucket in a multi-bucket pixel.

3.2.3 Shared Pixel Architecture

A key technique to enhance pixel shrinking in recent years is through the use of shared pixel architecture. In a shared architecture, readout circuitry including the RST, SF, and ROW transistors are shared between neighboring pixels [Nakamura 06]. Figure 3.5 shows one example in which two pixels in the neighboring rows share the same readout circuitry. As we have discussed in Section 3.1, the pixel transistors can be considered as an overhead because they reduce the pixel's fill factor. However, despite the fact that pixel size keeps shrinking [Agranov 09] [Agranov 11], transistors that constitute the readout circuit do not shrink as fast, or at all due to other concerns. For example, reducing the length of the source follower transistor causes an increase in its random telegraph signal (RTS) noise [Findlater 04] [Wang 06] [Leyris 06] that dominates read noise and degrades sensor performance. Therefore, a shared architecture is effective in maintaining the pixel's fill factor by reducing the effective number of transistor per pixel. In order to use a shared architecture, the location where photo-generated charges are accumulated must be isolated from the sensing node FD. In the 4T pixel, this requirement is fulfilled because the TX gate isolates the PPD and the FD.

For a multi-bucket pixel, being able to use shared pixel architecture is important because the pixel fundamentally has smaller fill factor compared with a conventional pixel as extra pixel areas are needed to accommodate the buckets.

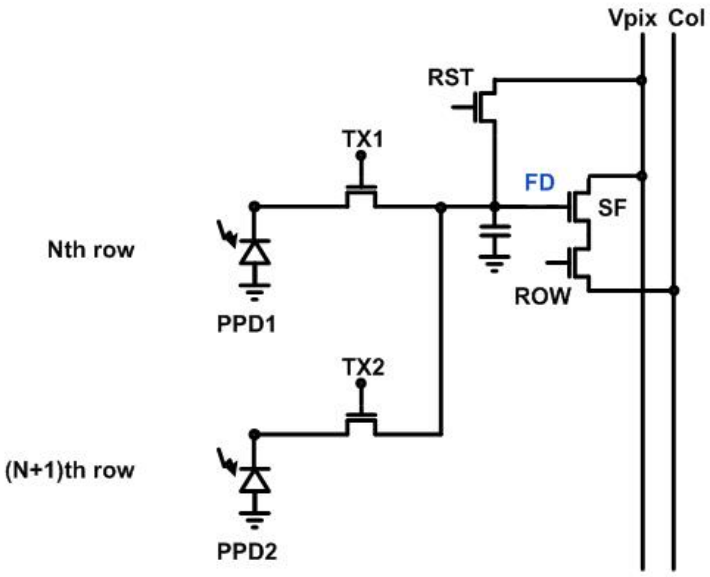


Figure 3.5: *A vertical two-way shared pixel architecture.*

3.2.4 Scalability

A desirable feature unique to multi-bucket pixels is the scalability of the number of buckets. As we have seen in Chapter 2, the larger the number of buckets, the more images that can be captured through time-multiplexed exposure. However, the size of a bucket and the number of signal line required to control the bucket limit the number of buckets for a given pixel size. Especially if a small pixel is desired, it is important to have a bucket design that is very area efficient.

3.3 Related Work

In this section, we will review two most commonly used bucket implementations in the literature, and evaluate them using the metrics described in the previous section. The first option is to use a floating diffusion as a bucket [Yamamoto 06] [Kawahito 07] [Kim 10] [Stoppa 10]. Figure 3.6 shows a single-bucket pixel that uses this approach. The advantages of this implementation are its simplicity and compactness because in principle no new process development is needed and no additional transistor or control line is required. However, this approach has several drawbacks. First, CDS cannot be used because resetting the FD to readout the reset level would erase the signal charges accumulated there. As a result, this pixel has a high noise floor. For example, the noise floor of the sensor reported in [Kim 10] is $80e^-$ with most of them coming from the kTC reset noise. Second, depending on the fabrication process, dark current generation rate at the FD can be 1000X higher than that of the PPD [Holst 07]. Third, shared architecture cannot be used to minimize pixel size as the charge accumulation and sensing regions are at the same location. Finally, as shown in Figure 3.8 (a), a separate readout circuitry is required for each bucket therefore limiting the scalability of the number of bucket.

Another approach to implement a bucket is to use another PPD, one that is shield from light, and additional transistor and control signal line [Yasutomi 10] as shown in Figure 3.7. In contrast to the previous approach, since the storage PPD is isolated from the FD by a TX gate, reset level of FD can be measured without destroying the accumulated charges therefore allowing CDS to be performed. For the same reason,

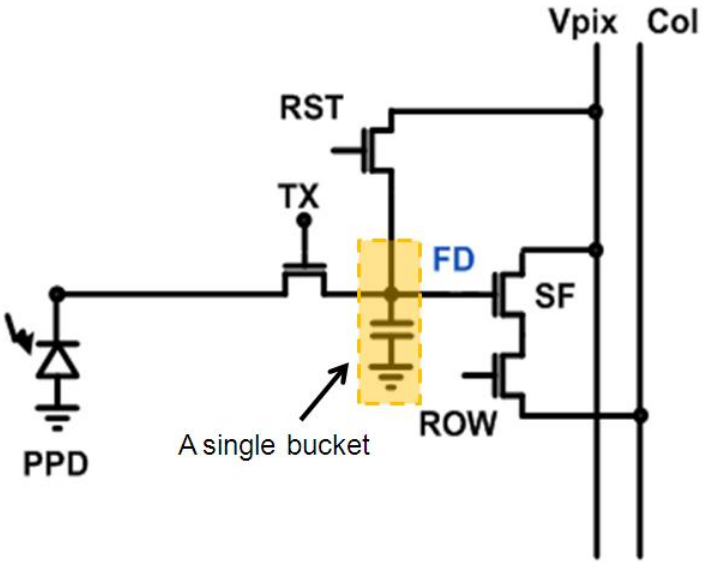


Figure 3.6: *Using FD as a bucket*

shared pixel architecture can be used to minimize pixel size and the same readout circuitry can be shared by all the buckets. Finally, dark current performance of this pixel is excellent because both the photodiode and the bucket are implemented using PPDs. The only drawback of this implementation, however, is that additional signal line and gate are required for each bucket added, causing a scalability issue as shown in Figure 3.8 (b).

3.4 Functional Analysis of a Bucket

In this section, we analyze a bucket from a functional perspective to understand the constraints and obtain insights on how a multi-bucket pixel should be designed. First, we would like the multi-bucket pixel to be able to cancel kTC noise through CDS.

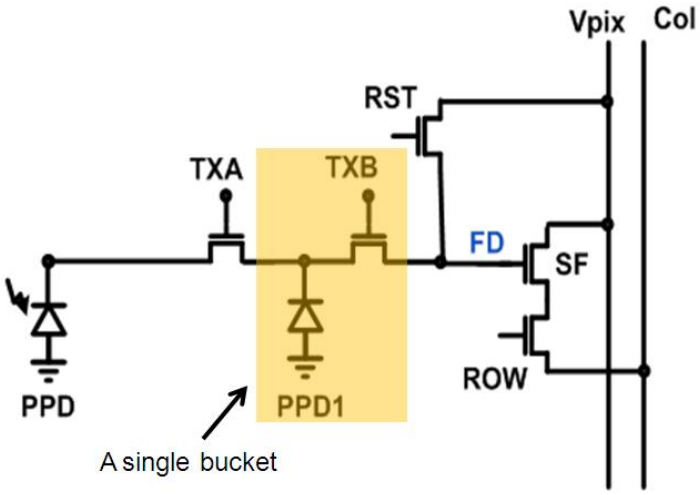


Figure 3.7: *Using PPD as a bucket*

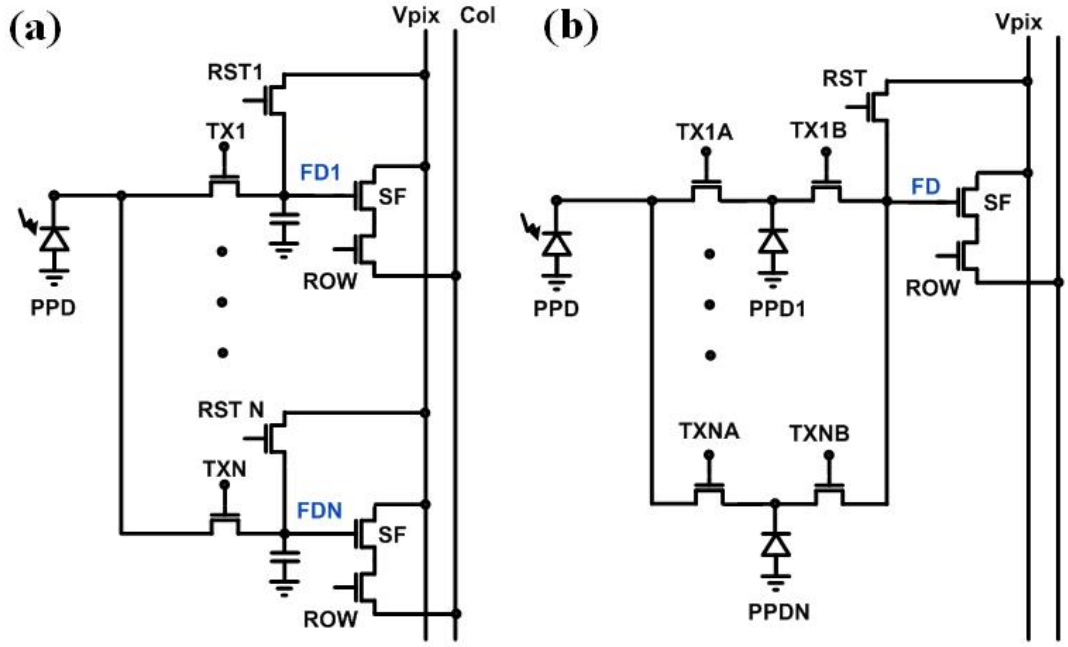


Figure 3.8: *A multi-bucket pixel using (a) FD and (b) PPD as in-pixel memories.*

Therefore, one constraint is that we cannot use the FD as a bucket. It means that we need a bucket elsewhere between the PPD and the FD as shown in Figure 3.9 (a). Second, we would like to be able to control which light goes into which bucket, therefore we need to have a switch between the PPD and the bucket as shown in Figure 3.9 (b). Finally, we need another switch between the bucket and the FD in order to do CDS to cancel kTC noise. As a result, we end up having a configuration as shown in Figure 3.9 (c). The pixel shown in Figure 3.8 (b) in fact fits into this configuration but it has the scalability issue mentioned above. Therefore, this research explored the possibility to come up with a single compact device that can implement the same function as the bulky structure shown in Figure 3.9 (d). A multi-bucket pixel which is compact, scalable, and preserves true CDS to cancel kTC noise simultaneously is described in the next section and requires a fundamentally different pixel design.

3.5 Proposed Pixel Structure and Operation

Using a single compact device to replace a bulky structure is not a totally new problem. Researchers encountered similar problems a long time ago in the CCD world. One close example is virtual phase CCD (VPCCD) invented in 1979 [Hynecek 79]. The goal of VPCCD is to replace one of the gate electrodes in a two-phase, buried channel CCD operating in the 1+1/2 phase mode by implant. Reduction of a gate electrode eliminates the possibility of gate-to-gate shorts encountered back then in other CCD technologies, and also has the added advantage of producing devices with high quantum efficiency and excellent uniformity [Hynecek 79]. In this section, we

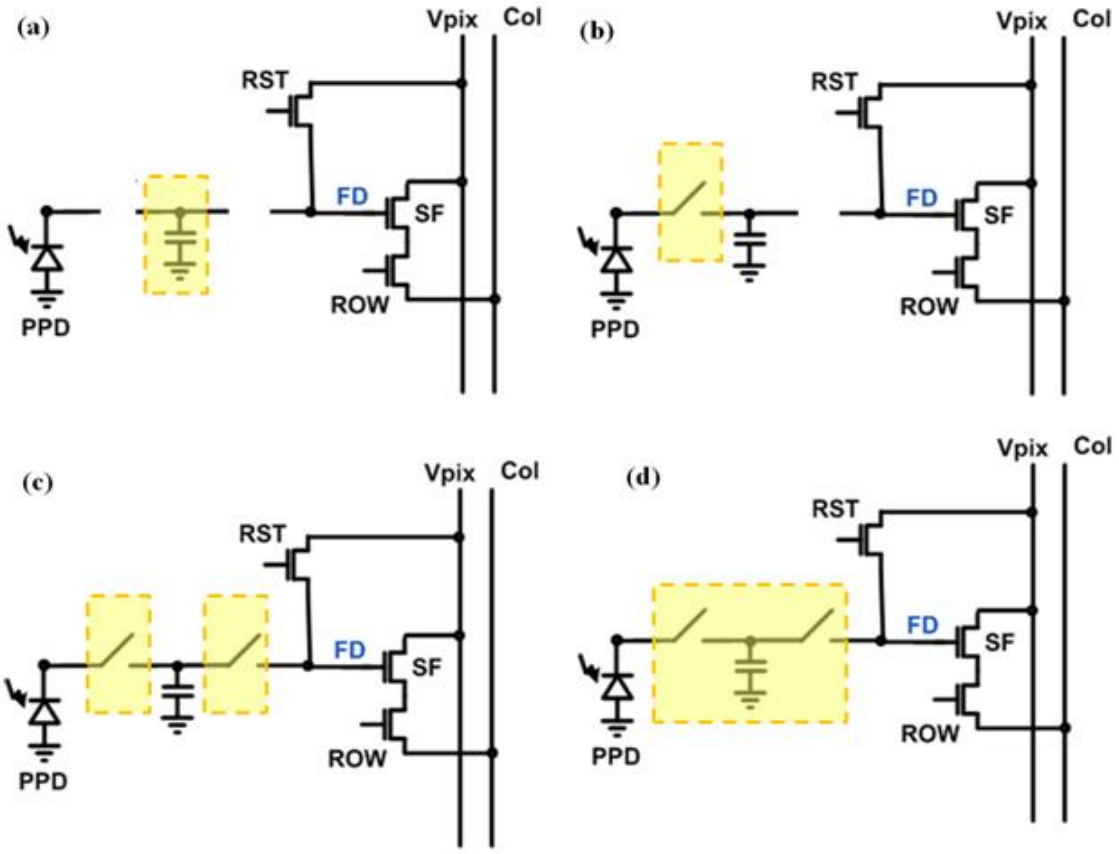


Figure 3.9: *Functional analysis of a bucket (a) The bucket needs to situate between the PPD and FD for CDS (b) A switch needs to exist between the PPD and the bucket to control which light goes into which bucket (c) Another switch needs to isolate the bucket and the FD for CDS (d) A bulky bucket to be replaced by a single compact device.*

described how the concept of VPCCD is incorporated into a conventional 4T CMOS pixel to design a new class of multi-bucket pixel.

Figure 3.10 shows a device structure which forms the basis of our multi-bucket pixel. It consists of a storage gate (SG) and a virtual barrier (VB) placed between a PPD and a FD. Similar to a VPCCD, a clocked barrier (CB) and a clocked well (CW) are formed under the SG by an additional n-type implant at the CW region while the VB region is formed by receiving one more surface p+ implant which is the same one used to form the PPD. These implants fix the potential of the VB and make the potential at the CW permanently higher than that of the CB. The potentials of CB and CW can be raised and lowered by application of appropriate voltages on the SG [Hynecek 79]. Since it is a non-standard device from CCD, it is challenging to build it in CMOS and we need to modify the existing CMOS image sensor fabrication process. The key challenge is getting the doping correct to create the needed potential barriers for the charges in the silicon. It involves coming up with correct implants including both the energy and dosage, and designing various masks that define the locations of these implants. In the whole design process, we have used TCAD tools such as Silvaco's ATHENA and ALTAS, and Sentaurus to guide the design. Figure 3.11 shows simulation results of the device's net doping profile.

Differently from VPCCD, we clock the SG with tri-level voltages such that the region circled by the black dashed line in Figure 3.10 replaces the bulky structure in Figure 3.9 (d) by acting as both transfer and storage device. Figure 3.12 shows 2D electrostatic potentials in the silicon and maximum potentials along the orange

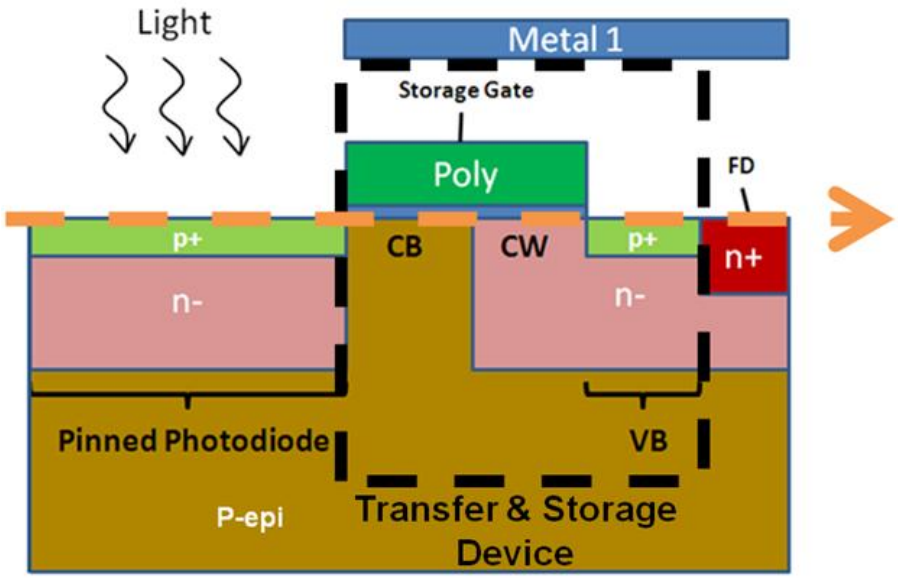


Figure 3.10: *Proposed device structure. The structure between the PPD and the FD acts as both transfer and storage device. The CW is formed by n- implant while the VB is formed by an additional surface p+ implant.*

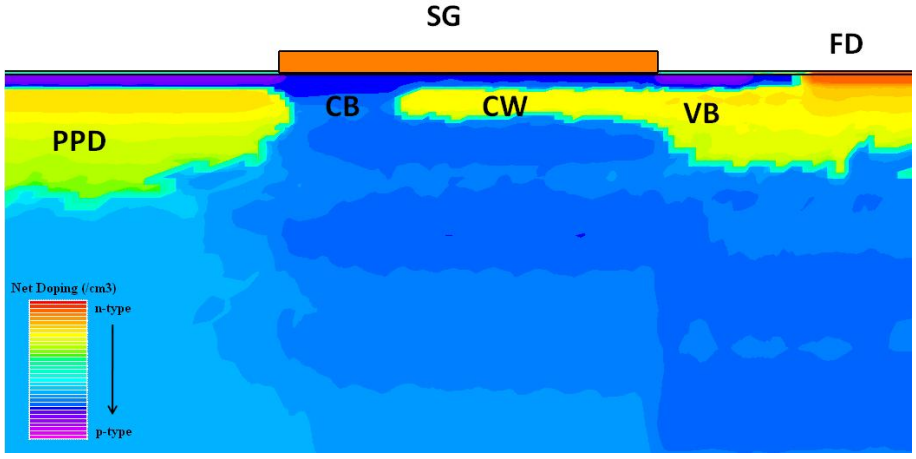


Figure 3.11: *Simulated net doping profile of the proposed device structure using Silvaco's ATHENA.*

line in Figure 3.10 under different SG biases to illustrate the device operation. When a high voltage is applied to the SG, potentials at CB and CW become higher than that of the PPD causing electrons to flow from the PPD to the CW. The potential at the VB is lower than that of the CW and it prevents electrons from further flowing to the FD. When a mid-level voltage is applied to the SG, the potential at the CW is higher than those at the CB and VB and therefore electrons previous transferred from the PPD are hold in the CW. The CW thus serves as an in-pixel memory. The doping profile of the channel is designed such that PPD and CW are isolated by the CB when the SG is biased at the mid-level. It allows a next packet of electrons to be accumulated in the PPD without mixing with electrons being stored in the CW in this phase. Additionally, the VB that isolates the CW and the FD in this phase means that the reset level of the FD can be measured thus allowing true CDS to be performed. Finally, when a low voltage is applied, the potential at the CW becomes lower than that of the VB and so electrons are transferred from the CW to the FD for readout.

As shown in Figure 3.13, by operating the SG using tri-level voltages, the proposed device structure can be used to replace the bulky structure shown in Figure 3.9 (d) by acting as both transfer and storage device therefore saving gate count, signal control lines, and allowing us to create pixel with a smaller pitch. Figure 3.14 shows a circuit schematic of a multi-bucket pixel that uses the proposed device structure. Compared with using a PPD as a bucket, this implementation requires a smaller area overhead since it eliminates one gate and one metal routing per bucket making it more

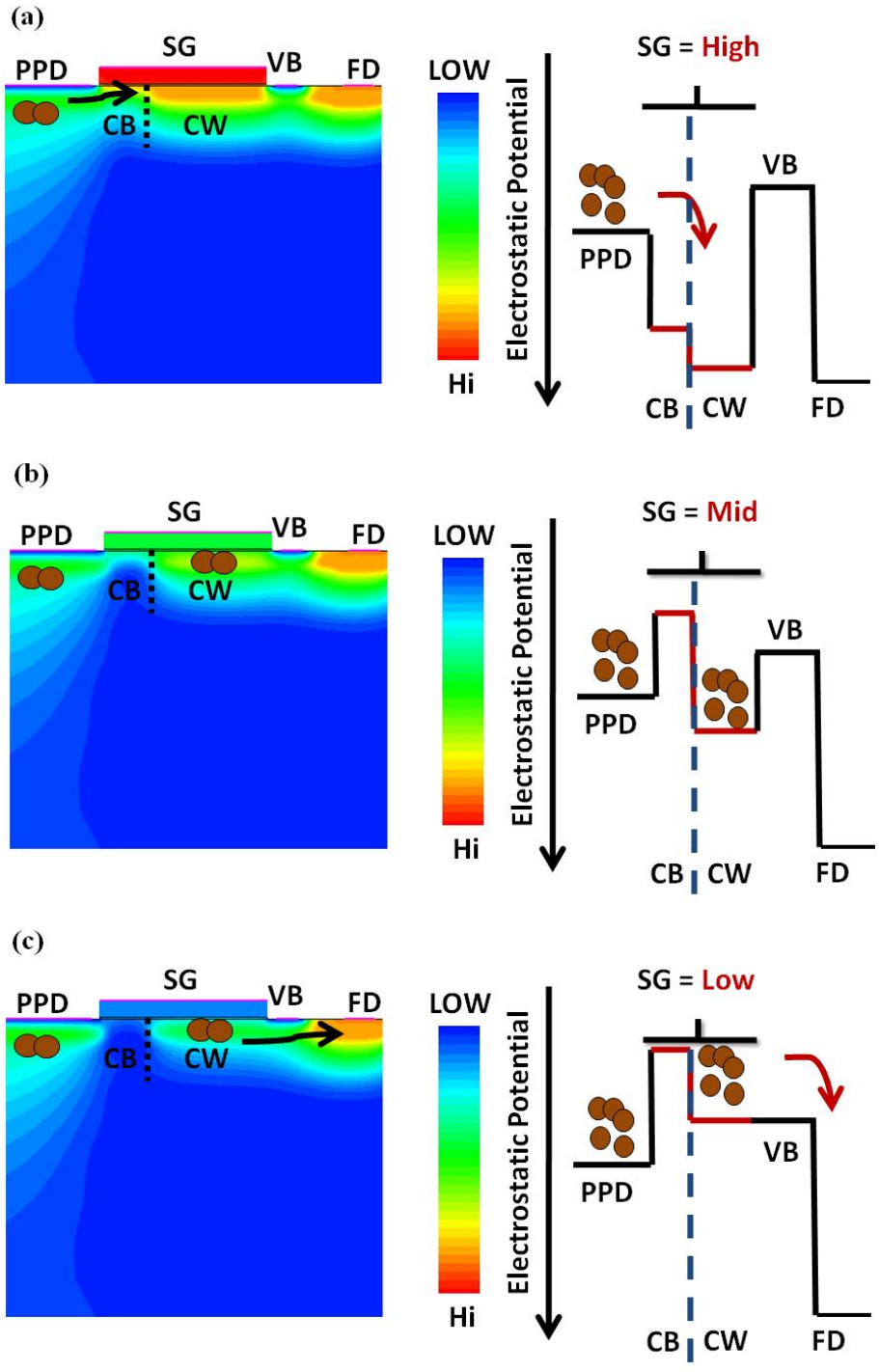


Figure 3.12: *Operation of the proposed device structure. (a) A high voltage is applied to SG to transfer charges from PPD to CW (b) A mid voltage is applied to SG to store charges under CW (c) A low voltage is applied to SG to transfer charges from CW to FD over VB.*

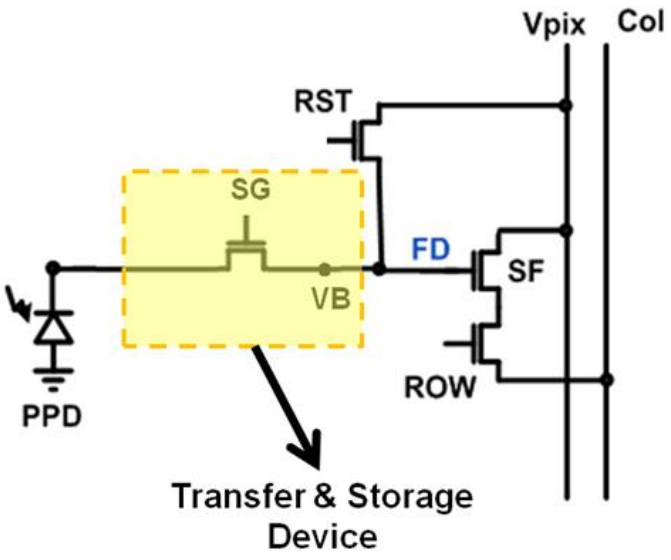


Figure 3.13: *Circuit schematic of a single compact bucket.*

compact and scalable. Compared with using a FD as a bucket, this implementation supports CDS to eliminate kTC noise. Additionally, the same readout circuitry can be shared by all the buckets and shared pixel architecture can be used to further reduce pixel size. This pixel structure has all the desirable features listed in Section 3.2 except that the bucket is not implemented using a PPD and therefore its dark current performance is a concern. We will discuss later in Section 4.4, without any optimization of the current fabrication process, this bucket in fact has a higher dark current than the PPD. Based on this pixel technology, we designed and fabricated two CMOS image sensors with dual and quad-bucket pixels which will be presented in the next chapter.

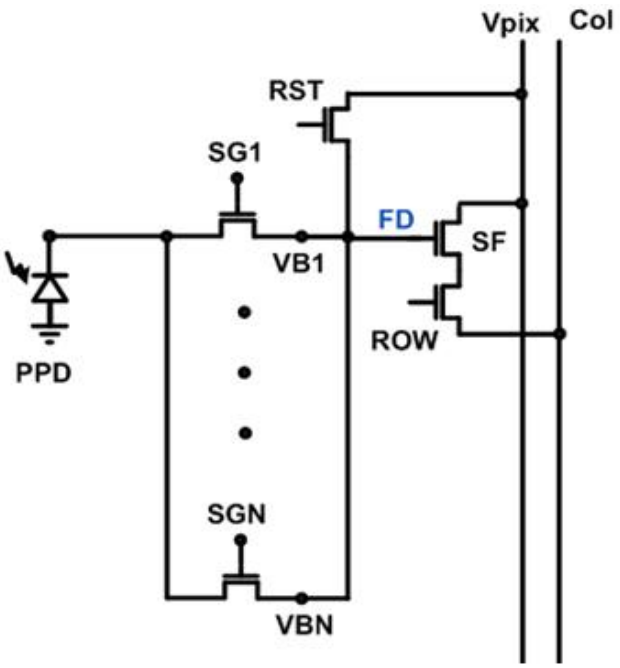


Figure 3.14: *A multi-bucket pixel using the proposed device structure.*

Chapter 4

CMOS Image Sensors with Multi-Bucket Pixels

In this chapter, we present the design and characterization of two CMOS multi-bucket sensors with dual and quad-bucket pixels. Using the pixel technology described in Chapter 3, sizes of the two pixels are the smallest among similar pixels reported in the literature. We first describe the pixel architecture, layout, and operation of each sensor. Then, we describe our characterization system and present experimental results of the two sensors.

4.1 Dual-Bucket Image Sensor

The dual-bucket image sensor comprises a 640_Hx576_V array of 5.0μm pixels in 0.11μm CMOS technology. Figure 4.1 shows architecture and conceptual layout of the pixel.

Two FDs of the same pixel are connected to the same readout circuit and a vertical two-way shared architecture is used. As explained in Chapter 3, such a shared architecture is possible because our pixel technology does not use a floating diffusion (FD) as a bucket. Two storage gates (SG) and an anti-blooming (AB) gate are connected to a pinned photodiode (PPD). Each of the SGs and its associated virtual barrier (VB) resembles the device structure described in the previous chapter and they serve as the two buckets of the pixel. The AB gate connected to the PPD is used for anti-blooming protection and resetting the PPD. Upper layer metals are routed to cover the buckets. They act as light shields to minimize parasitic light from contaminating signal charges being stored in the buckets.

By using the shared architecture, this pixel has a good fill factor of 42%. One drawback of using the shared architecture is that it increases the total capacitance of the FD (C_{FD}), therefore lowers the pixel's conversion gain and increases the read noise of the sensor as explained in Section 2.6.1 and Section 3.1. The reason for an increase in C_{FD} is that a long piece of metal is required to connect FDs of the two pixels together to the readout circuit. As shown in Figure 4.2, electrostatic couplings from neighboring signal lines increase the metal parasitic capacitances of the FD metal. Additionally, the diffusion areas double and therefore the device component of the C_{FD} also increases. Given this architecture, we optimized the routings of the signal lines to lower the C_{FD} as much as possible and by using Silicon Frontlines F3D tool [Ershov 11], we can simulate this FD capacitance and therefore predict the pixel's conversion gain very accurately as we will see in Section 4.4.

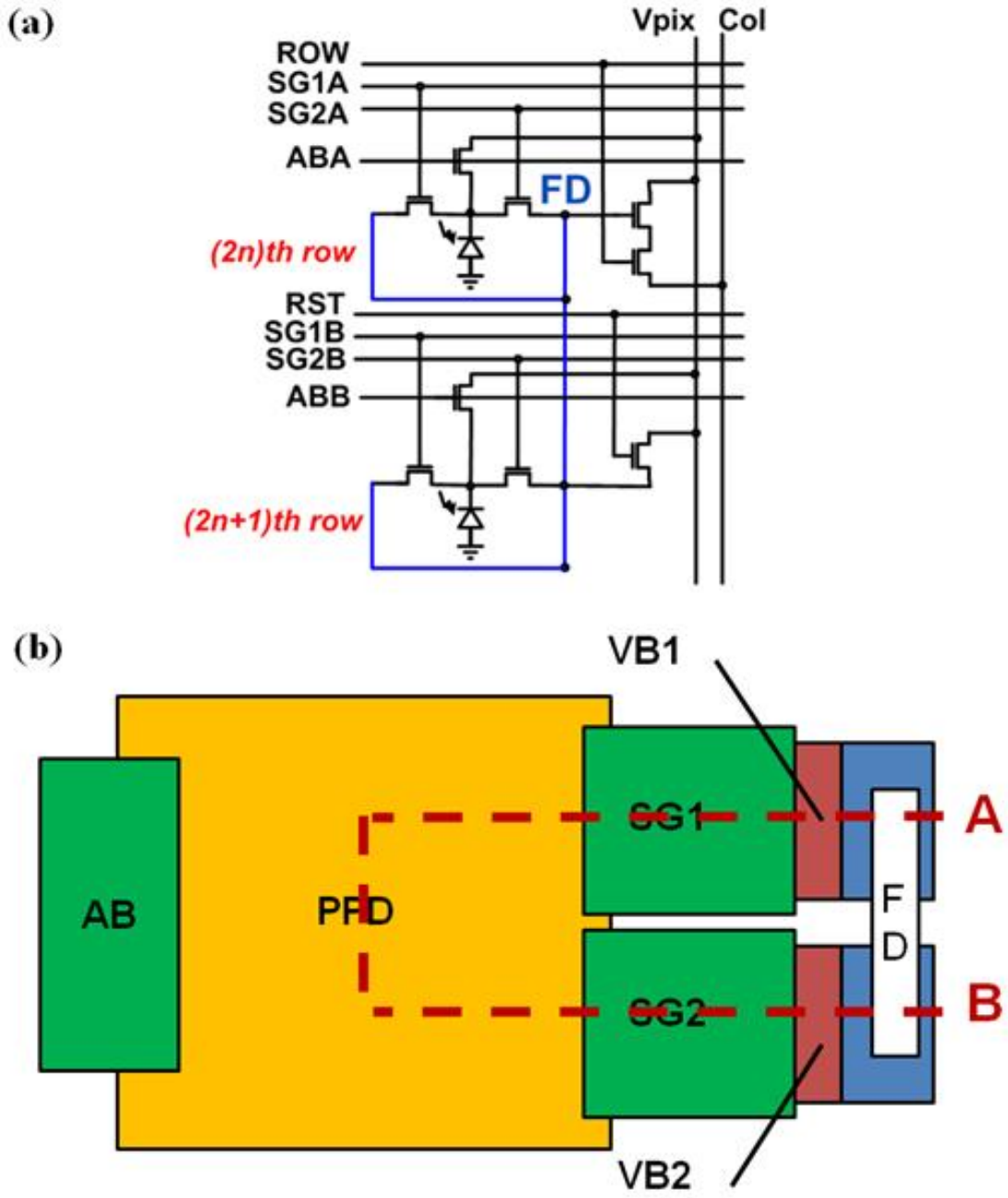


Figure 4.1: *Dual-bucket pixel. (a) Two-way shared architecture (b) Conceptual layout.*

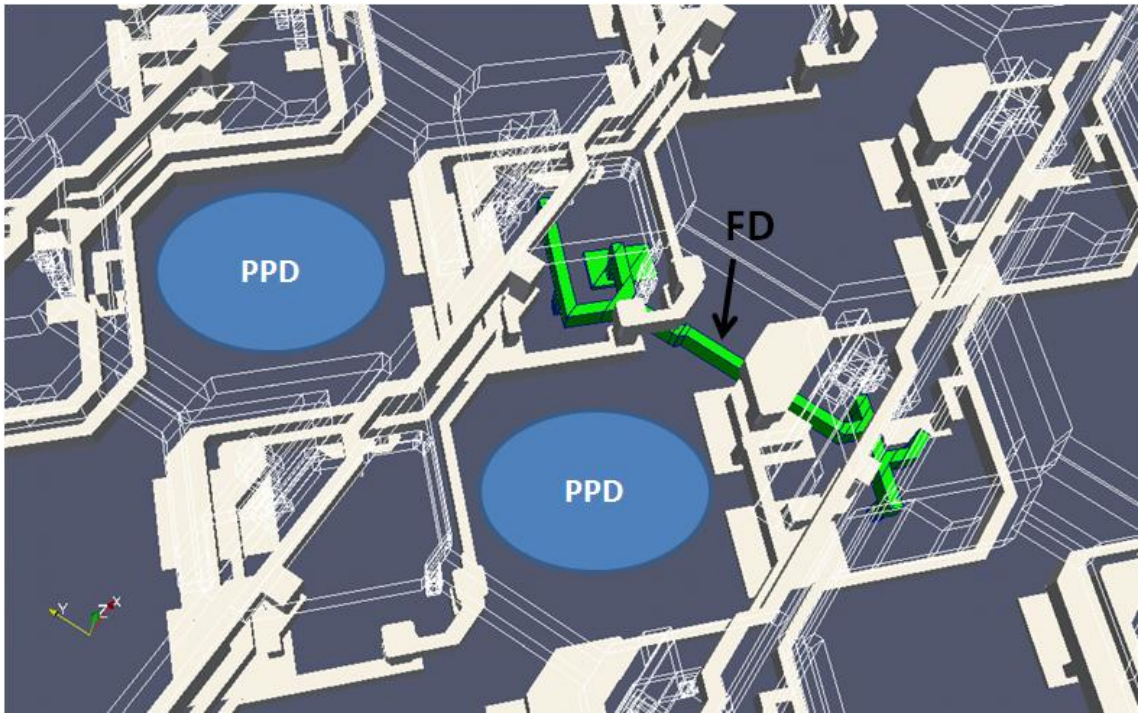


Figure 4.2: *Three-dimensional view of the dual-bucket pixel layout. The green line represents the metal used to connect the two composite floating diffusions of the two pixels. For clarity, some signal lines are omitted in this figure.*

Since the sensor has two buckets per pixel, it can be used to capture two time-interleaved images corresponding to two exposure conditions. Figure 4.3 shows a typical timing diagram of the sensor. Figure 4.4 shows the corresponding simulations of potential profiles using Silvacos ATHENA and ATLAS. During time-multiplexed exposure, the two SGs toggle synchronously with the exposure conditions between high (e.g. 4V) and mid-level voltages (e.g. 0.6V-1.6V) to transfer photo-generated electrons to the respective clock well (CW) under the SG where charges acquired under the same condition accumulate. At the end of the exposure, the same row is addressed twice to readout the signals being stored in the two SGs sequentially such that two interlaced frames are output by the sensor in every frame period. In each read, the FD is reset and the reset level is sampled. The VBs that separate CWs and the FD prevent this reset action from affecting the charges that are being stored in the CWs. To complete the true CDS operation, a low voltage (e.g. -0.5V) is applied to the SG to transfer charge from its CW to the FD.

4.2 Quad-Bucket Image Sensor

To explore the design challenges of scaling up the number of buckets per pixel, and to enable computational photography techniques that require capturing more than two exposures, a next-generation multi-bucket image sensor was designed and fabricated. This sensor has four buckets per pixel and contains an array of 640_Hx512_V 5.6μm pixels in 0.13μm CMOS technology. This design doubled the number of buckets per pixel, improves pixel capacity, and maintains a similar size of the dual-bucket pixel

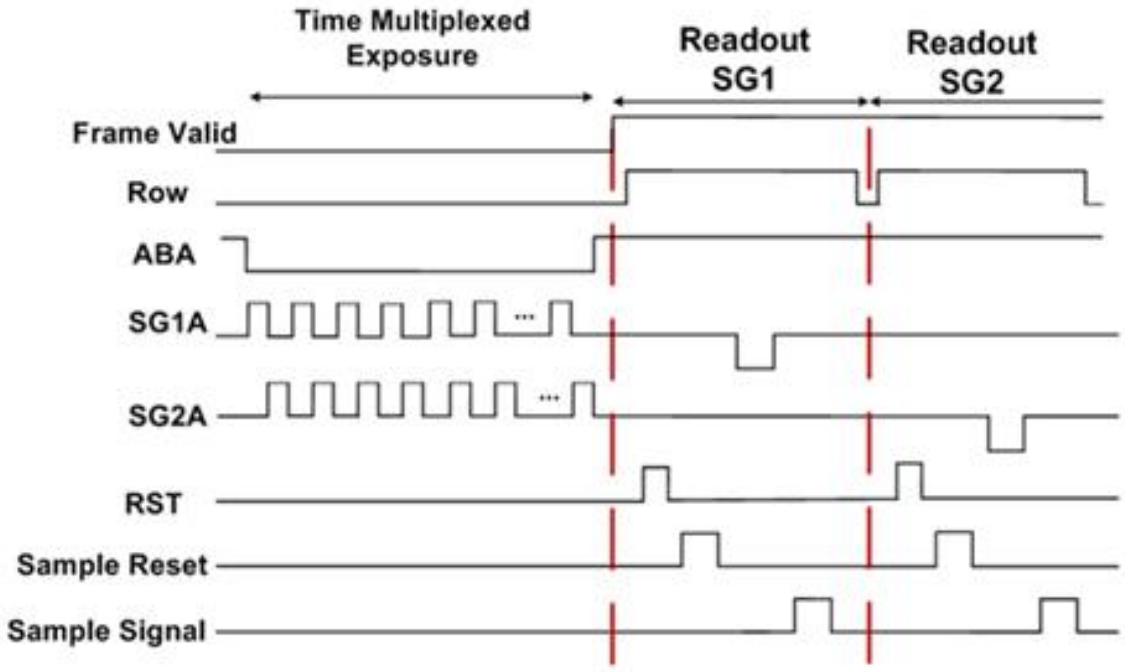


Figure 4.3: *Typical timing diagram of the dual-bucket sensor. The two SGs (SG1A and SG2A) can toggle arbitrarily during time-multiplexed exposure depending on application. The pattern shown here is one of the examples.*

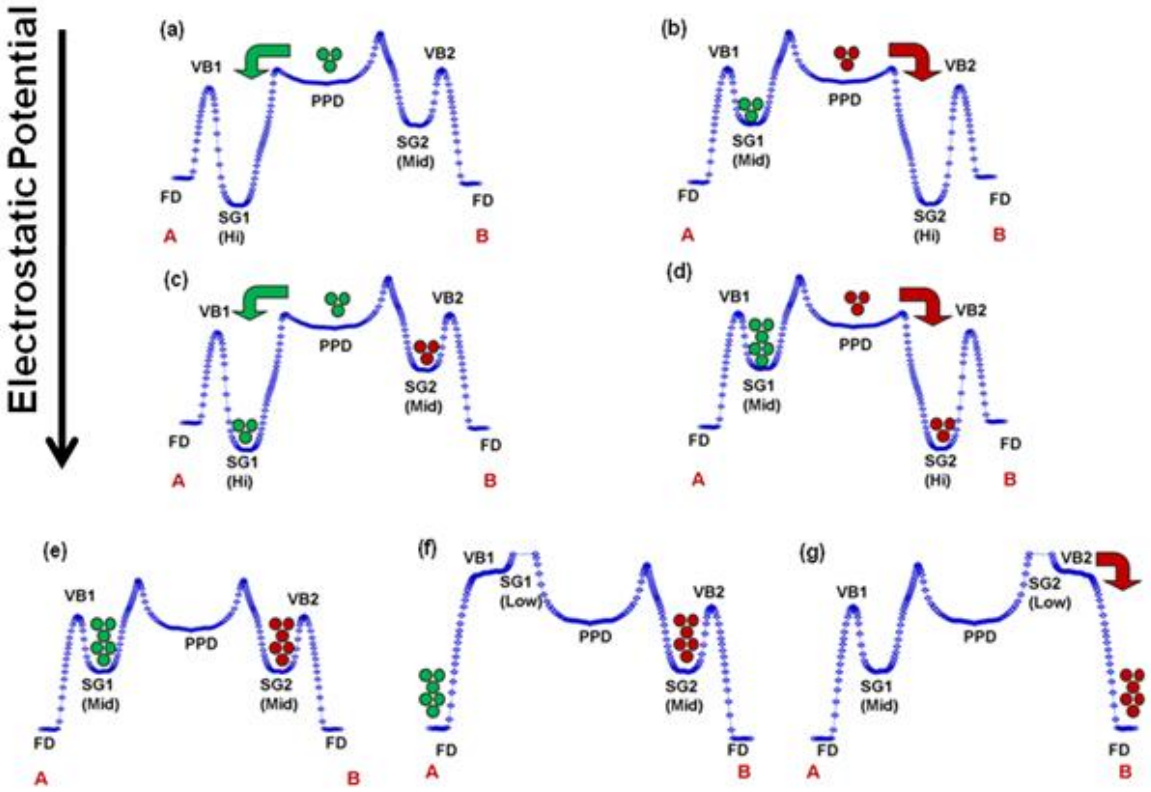


Figure 4.4: *TCAD simulations of pixel potential profiles along line A-B during time-multiplexed exposure (a) Charge transfer from PPD to SG1 (b) Charge transfer from PPD to SG2 (c) Second charge transfer from PPD to SG1 (d) Second charge transfer from PPD to SG2 (e) Charges stored under SG1 and SG2 (f) Charge transfer from SG1 to FD (g) Charge transfer from SG2 to FD.*

in terms of lithographic features.

Since the quad-bucket pixel needs to accommodate two extra buckets, an immediate design challenge is how pixel performance can be maintained or even improved without area overhead. Figure 4.5 shows circuit schematic and conceptual layout of the pixel. Four SGs and two AB gates are connected to a PPD. Again, the AB gates serve to reset the PPD and provide anti-blooming protection. The SG and its associated VB resembles the device structure described in the previous chapter and they serve as the four buckets of the pixel. Again, all the buckets are covered by upper layer metals to protect them from parasitic light. To improve optical and electrical symmetries, the four SGs are placed at the four corners of the PPD so that charges generated at the PPD would take approximately the same amount of time to travel to each of the CWs under the SGs.

This design connects the four FDs at the corners together by metal as shown in Figure 4.6. For simplicity, we represent each SG as a switch in the figure. Connecting all the FDs of the same pixel together allows them to share the same readout circuitry but it lowers the pixel's conversion gain due to the large parasitic capacitances of the long FD metal routing. Compared with the dual-bucket pixel which also has the FDs connected together, the impact on conversion gain in the quad-bucket pixel is more severe because the FDs are farther apart at the corners of the pixel and therefore the metal required to connect them is longer. Furthermore, since there are two more FDs per pixel, the diffusion areas and therefore device capacitances are larger in this pixel compared with that of the dual-bucket one. As discussed in the previous section,

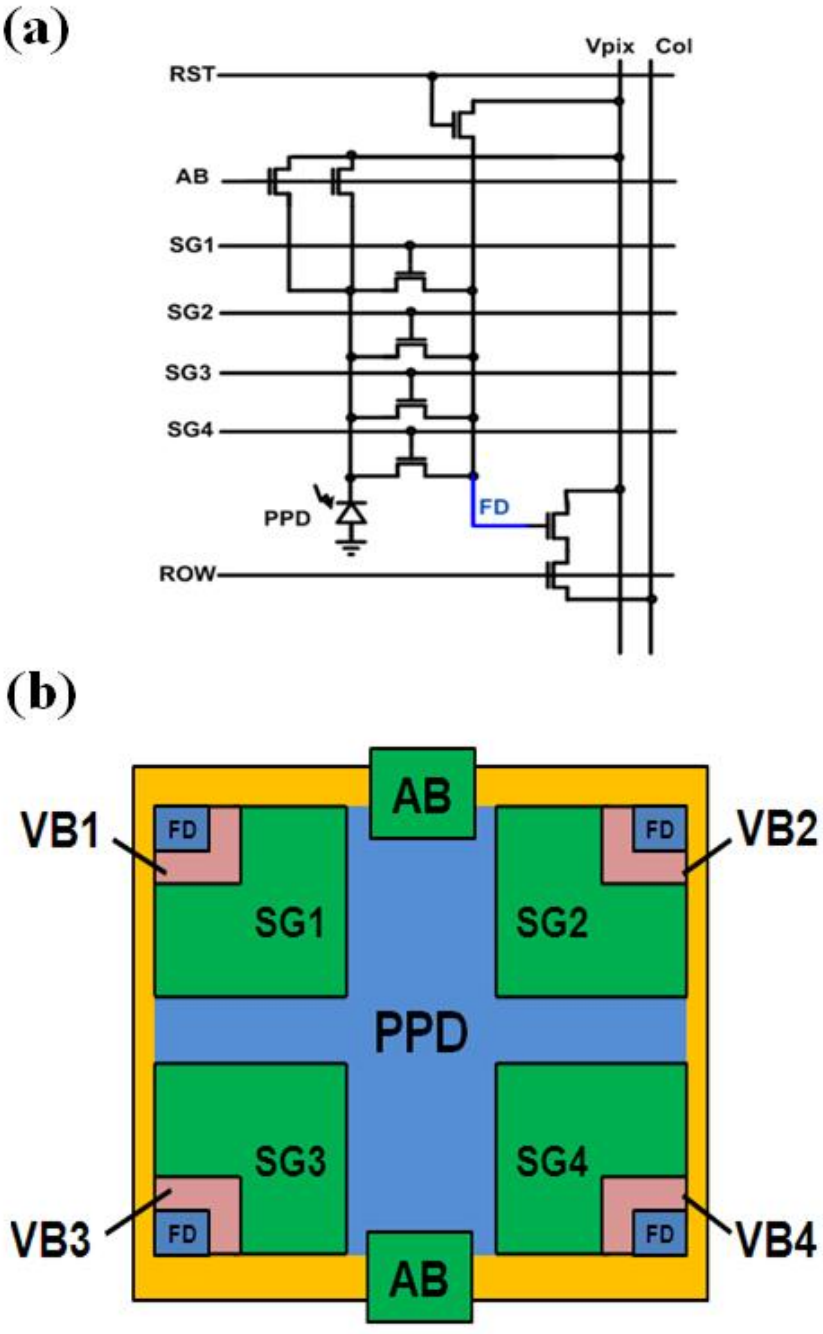


Figure 4.5: *Quad-Bucket Pixel. (a) Circuit schematic (b) Conceptual layout.*

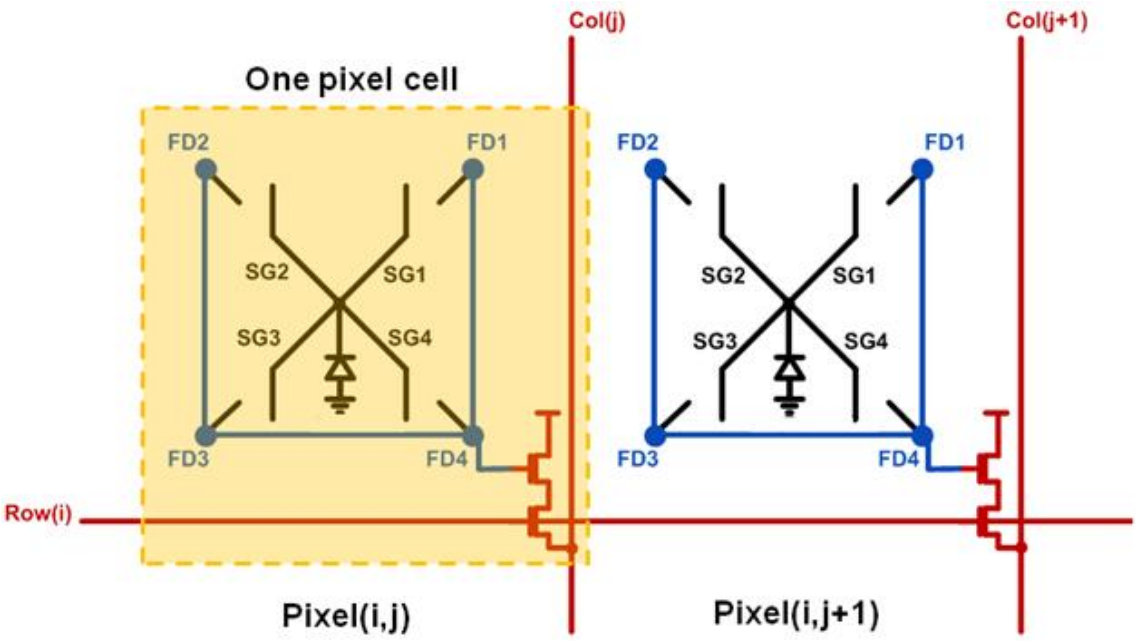


Figure 4.6: *Unshared FD quad-bucket pixel architecture. The current quad-bucket pixel connects FDs of the same pixel together by a long metal. The blue line represents a metal routing connecting FDs and the gate of source follower.*

using a shared architecture would normally increase the total FD capacitance and further drop the conversion gain. Due to this concern, the quad-bucket pixel uses an unshared architecture. Figure 4.7 shows a three-dimensional view of the pixel layout.

After the design was completed, we come up with an alternative design that shares FDs efficiently so that the pixel's conversion gain can be improved. We describe below this alternative design that we have not implemented in silicon. To create such architecture, first the long piece of metal that connects all the FDs of the same pixel together needs to be removed. But each FD now still needs to be connected to a readout circuit. To minimize metal parasitic capacitance, we can connect the nearest FDs together even though they belong to different pixels, and use the same readout

Figure 4.7: *Three-dimensional view of the quad-bucket pixel layout. The green line represents the metal used to connect the four floating diffusions of the pixel. For clarity, some signal lines are omitted in this figure.*

circuit for them as shown in Figure 4.8(a). To minimize device capacitance due to the diffusion area, we can further merge the FDs together to form a composite region such that no metal is needed to connect the FDs while only a small piece of metal is required to connect the composite FD to the readout circuit as shown in Figure 4.8(b).

This alternative pixel architecture shown again in Figure 4.9 not only improves conversion gain, but can also reduce the pixel size by tightly grouping the nearest FDs. Drawbacks of this implementation, however, would be a slightly more complicated readout because signals from the four SGs of the same pixel would be readout from different circuitries and column lines. In the example shown in Figure 4.9, image captured by SG0 and SG1 would be shifted one column to the right from those captured by SG2 and SG3. However, the peripheral digital logics can easily handle this situation. Another disadvantage of this alternative architecture is that it does not allow summing signals from the same pixel's SGs at the FD. When time-multiplexed exposure is not needed (e.g. when we want to capture one image only), a multi-bucket sensor can be turned into a single-bucket mode to increase total pixel capacity. It can be done by accumulating charges equally in the four buckets (e.g. through turning on the SGs simultaneously) during an exposure and subsequently summing them at the FD upon pixel readout. Since signals are summed in the charge domain, which can be a noiseless process, before they are readout, the image only suffers from one read noise. However, if such FD summing is not allowed, we need to readout four images and sum them in the voltage domain. As discussed in Section 2.6.1, the final

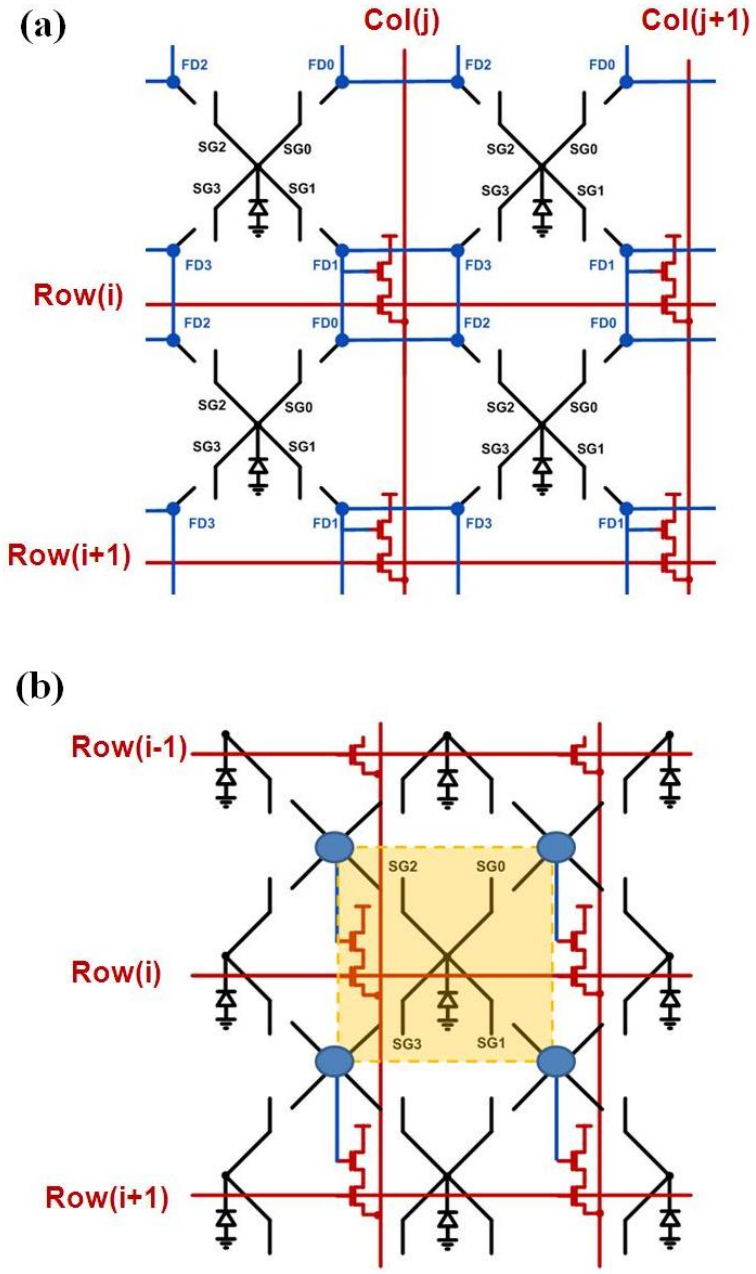


Figure 4.8: *Procedures to develop alternative quad-bucket pixel architecture with higher conversion gain. (a) Connecting neighboring FDs to minimize metal parasitic capacitance (b) Merge FDs to form a composite FD region to minimize device capacitance.*

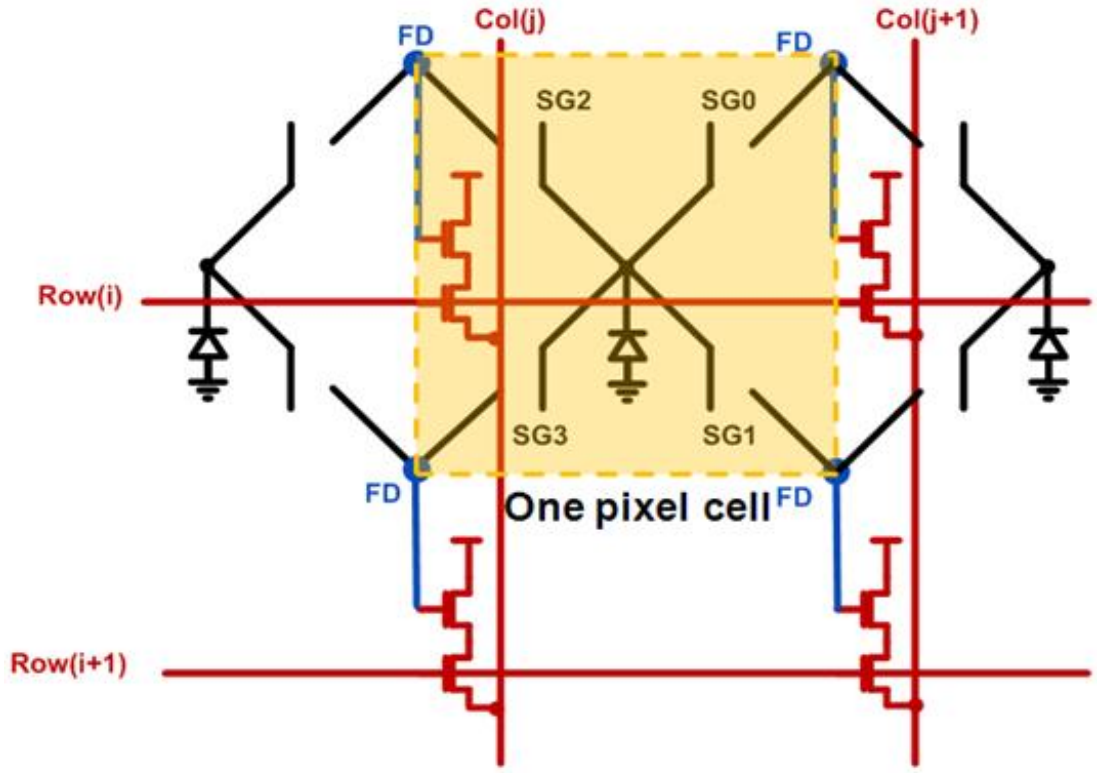


Figure 4.9: *Shared FD quad-bucket pixel architecture. Four nearest FDs from neighboring pixels are merged and the composite FD shares the same readout circuit.*

image suffers more read noise in this case. Besides, it requires 4 times more readout therefore incurring slower effective frame rate and more power consumption.

Since photo-generated charges accumulate in the PPD for only a short time before being transferred to a SG during time-multiplexed exposure, from a total charge capacity perspective we can enlarge the size of each SG and reduce that of the PPD to minimize the impact of adding two extra buckets in pixel area to keep pixel size small, comparable to the dual-bucket pixel. Doing so reduces the pixel's fill factor to 10% which causes a concern about its quantum efficiency. The effect of small PPD

size, especially for a large pixel, can be reduced through microlens and optical stack optimization such as adding a light guide [Hsu 04]. In this work, we simply used a standard microlens and did not attempt to optimize our light path. While this did affect our sensitivity (see Section 4.4), even a simple microlens was able to reduce the effect of the small fill factor.

Finally, charge transfer is improved by lowering the CBs (i.e. higher potential) in the quad-bucket pixel such that complete charge transfer from PPD to CWs occurs at a lower voltage (e.g. from 4 to 3.5V) than that in the dual-bucket pixel. Again, since charges are accumulated in the CWs instead of the PPD, such a change does not reduce pixel capacity even though PPD capacity is reduced.

Figure 4.10 shows a typical timing of the quad-bucket sensor. Similar to that of the dual-bucket sensor, the four SGs can toggle between high (e.g. 3.5V) and mid-level voltages (e.g. 0.6-1.2V) during time-multiplexed exposure while a low voltage (e.g. -0.5V) is used to readout the charges. The same row is addressed four times sequentially to readout the captured images in the four SGs using true CDS.

4.3 Characterization System

Figure 4.11 shows a simplified block diagram of the system used to characterize the sensors. To maximize programmability, all the timing signals including row and column addresses are provided by a FPGA while voltages and currents are supplied by a digital-to-analog converter (DAC) external to the sensor. An analog-to-digital converter (ADC) converts analog pixel values from the sensor to digital and sends the

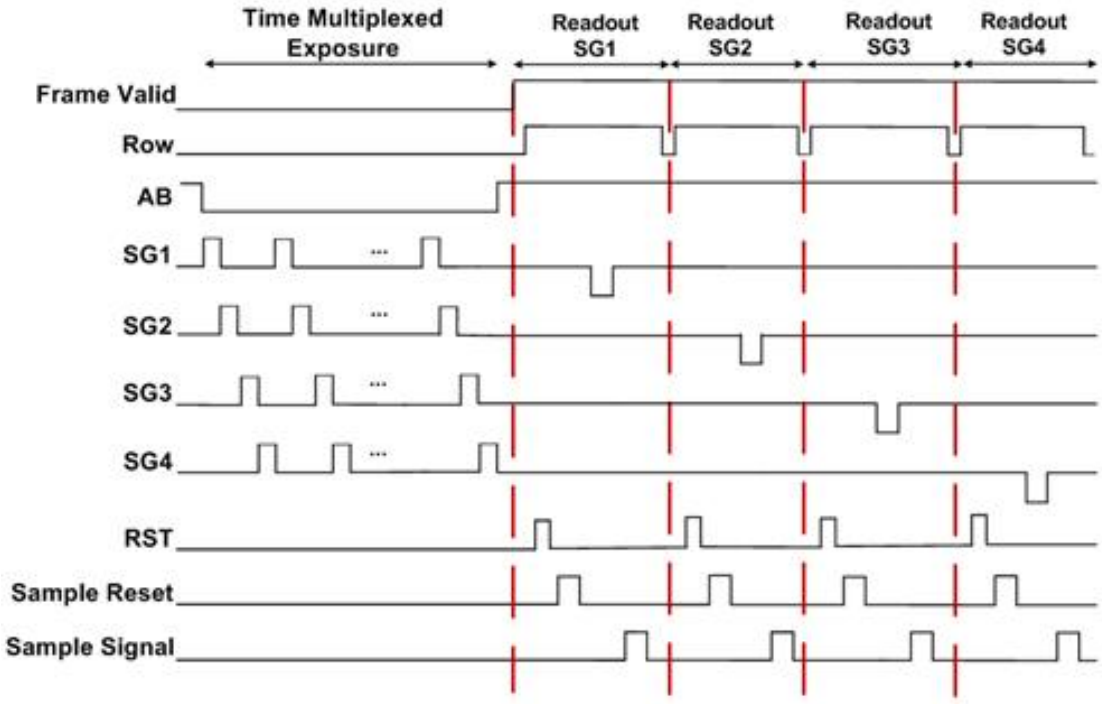


Figure 4.10: *Typical timing diagram of the quad-bucket sensor. The four SGs (SG1, SG2, SG3, and SG4) can toggle arbitrarily during time-multiplexed exposure depending on application. The pattern shown here is one of the examples*

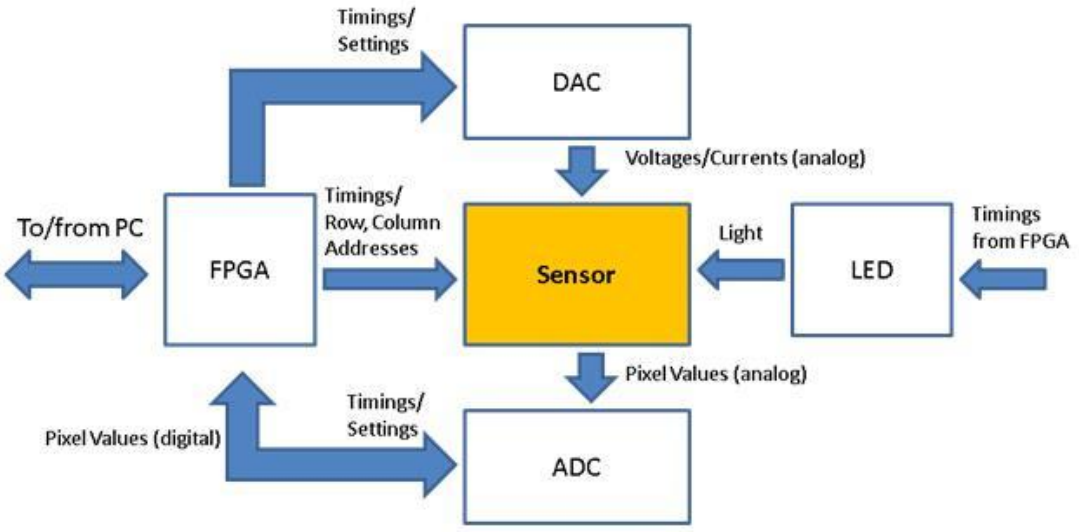


Figure 4.11: *A simplified block diagram of the system used to characterize the sensors.*

results to the FPGA which transfers them to a PC for display. The flexibility offered by the FPGA allows us to have complete control of the pixel array operation. On one hand, the flexibility allows us to experiment with and determine the optimal timings and voltages for pixel operation. On the other hand, new sensor operations such as flash/no-flash, multi-flash, HDR, etc can easily be implemented and experimented in this system. The LED, also being controlled by the FPGA, enables us to perform experiments involving synchronized active illumination.

4.4 Experimental Results

Figure 4.12 shows the chip micrograph and packaged die of the sensors. Besides the pixel array, both sensors contain similar peripheral circuitries such as row driver, column decoder, CDS circuitries, and a programmable gain amplifier.

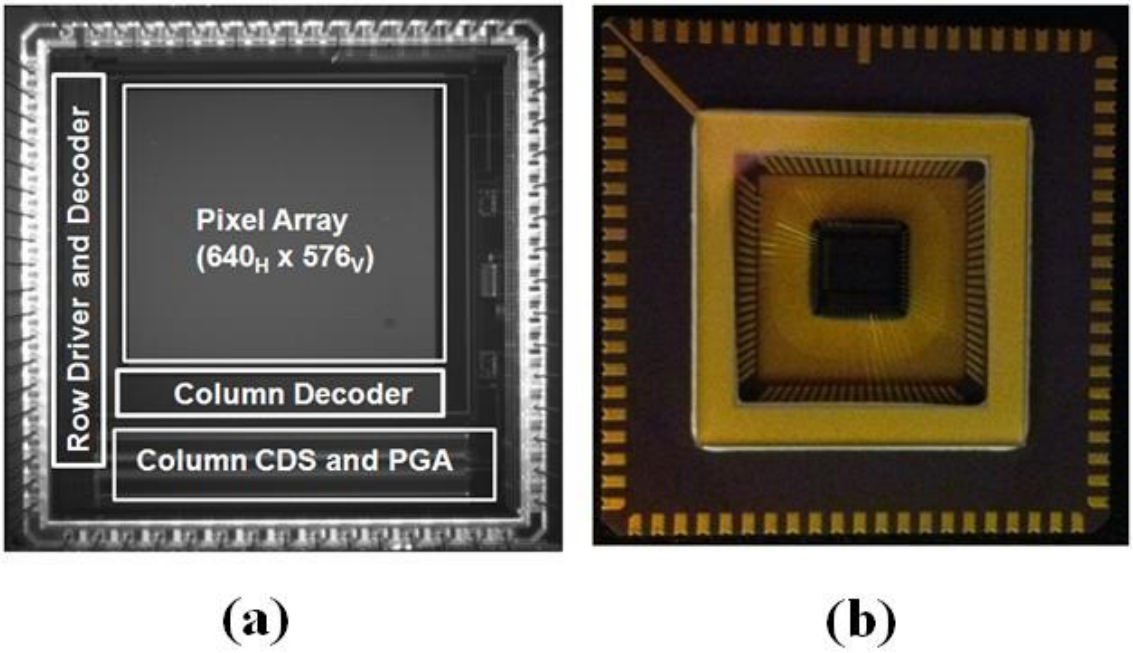


Figure 4.12: *(a) Chip micrograph of the dual-bucket sensor and (b) packaged die of the quad-bucket sensor*

	Dual-Bucket	Quad-Bucket
General Specifications		
Process	0.11 µm 2P4M	0.13 µm 2P3M
Chip Size	5.1x5.1mm^2	5.1x5.1mm^2
Power Supply (Analog/Digital)	3.3V	3.3V
Pixel Size	5µm	5.6µm
Pixel Architecture	2-Way Shared	Unshared
Fill Factor	42%	10%
Resolution	640_Hx576_V	640_Hx512_V
Number of Storage Node	2	4
Shutter Type	Global Shutter	Global Shutter
Sensor Characteristics		
Conversion Gain	51uV/e	33uV/e
Storage Node Capacity (@1.2V)	10ke	23ke
Read Noise (@Gain = 8X)	5.5e	8e
Dark Current (@25C)	20nA/cm^2	7.2nA/cm^2
Peak Quantum Efficiency	71.5%	39.5%
Responsivity (@550nm)	74ke/lux-s	49ke/lux-s
FPN (@50% Signal)	0.98%	1.10%
Temporal Noise (@50% Signal)	1.36%	0.93%

Table 4.1: *Summary of sensor characteristics at 1.2V SG mid-level voltage.*

Table 4.1 summarizes characteristics of the fabricated sensors. As we have predicted, the dual-bucket sensor has a higher conversion gain compared with the quad-bucket one because the later fundamentally needs more control signals and therefore metal routings which contribute to the FD parasitic capacitances through metal-to-metal coupling. However, as we have discussed in the previous section, the conversion gain of the quad-bucket pixel can be improved by using the alternative shared-FD pixel architecture. Simulations show that this alternative architecture can potentially achieve 1.8 times improvement in conversion gain as shown in Figure 4.13. A consequence of having a lower conversion gain is that the quad-bucket pixel has a higher read noise than that of the dual bucket one as shown in Figure 4.12.

Figure 4.15 shows the measured quantum efficiencies of the two pixels. As we have predicted, even though the quad-bucket pixel has four times smaller fill factor and no optimization has been made on the microlens and light path, its quantum efficiency is decreased by less than two times from the dual-bucket pixel across visible wavelengths. It indicates that even a simple microlens is effective in focusing light from the pixel opening to the PPD.

Figure 4.16 shows measured photo-responses of the two pixels with respect to different SG mid-level voltages. Storage capacity of a SG increases with increasing SG mid voltage. At 1.2V, the dual-bucket pixel has around 10ke^- per bucket while the quad-bucket pixel has around 23ke^- per bucket. Therefore, the quad-bucket pixel achieves more than 2X improvement in storage capacity per bucket under the same SG biasing voltage. Together with the fact that it has two extra buckets, the

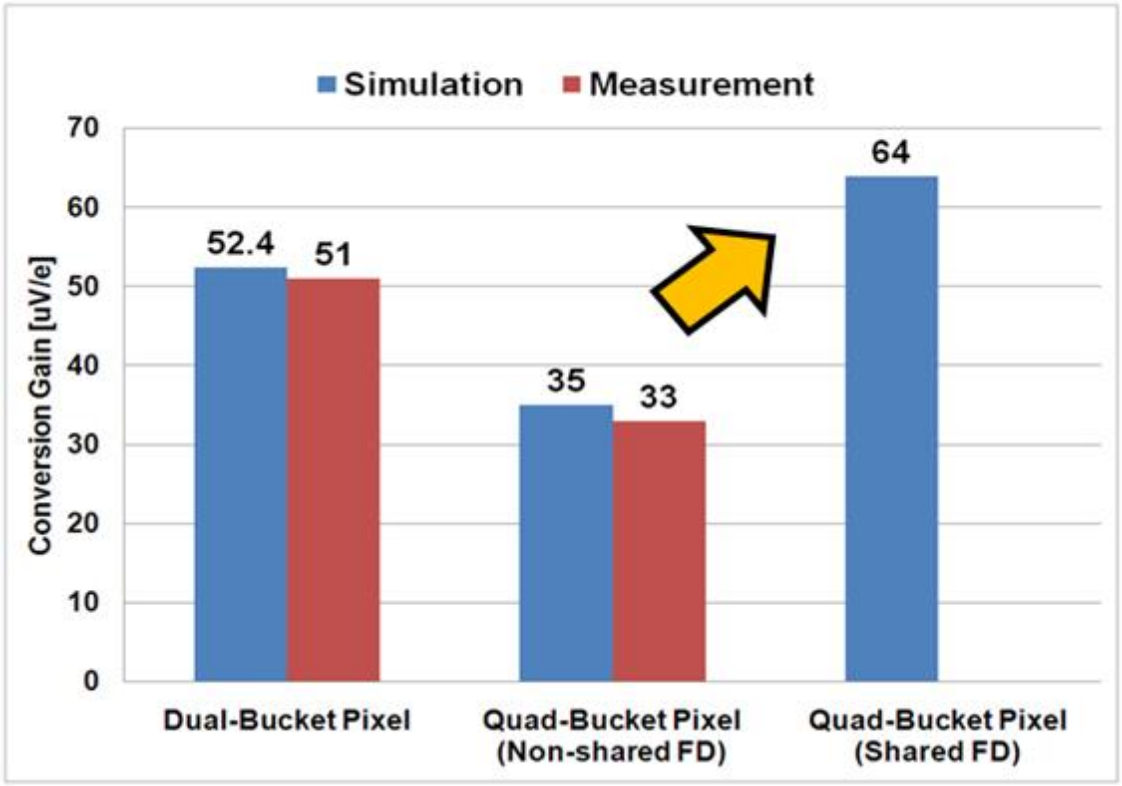


Figure 4.13: *Conversion gain simulation and measurement results.*

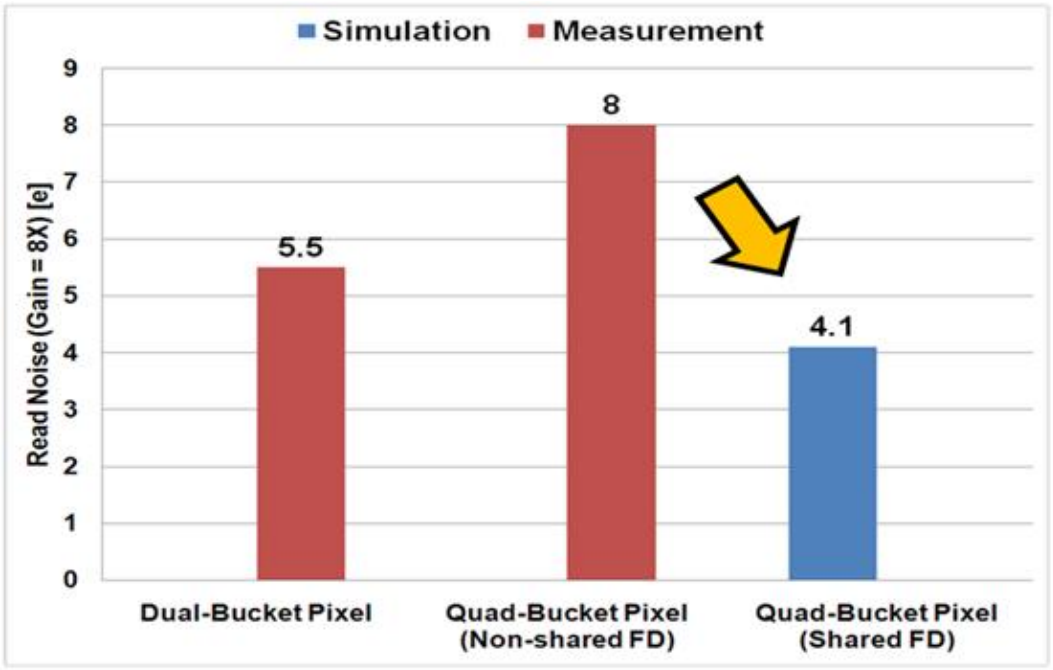


Figure 4.14: *Read noise measurement and simulation results.*

quad-bucket pixel has almost 90ke$^-$ total pixel capacity compared with 20ke^- in the dual-bucket pixel, i.e. more than 4 times bigger despite a slight increase in pixel size.

Figure 4.17 shows the measured SG dark currents of the two pixels again under different SG mid-level voltages at room temperature. Currently, the dark current is higher than that of state-of-the-art 4T imager pixels. But it should be noted that the dark current generation rate depends heavily on the fabrication process as we can see by comparing the dark currents of the dual and quad-bucket pixels made using different processes. No optimization has been made in tuning the processes in this work due to our limited control on the foundry. Nevertheless, despite we expect dark current would improve through process optimization, this pixel would fundamentally have worse dark current than if a pinned photodiode is used as a bucket because of

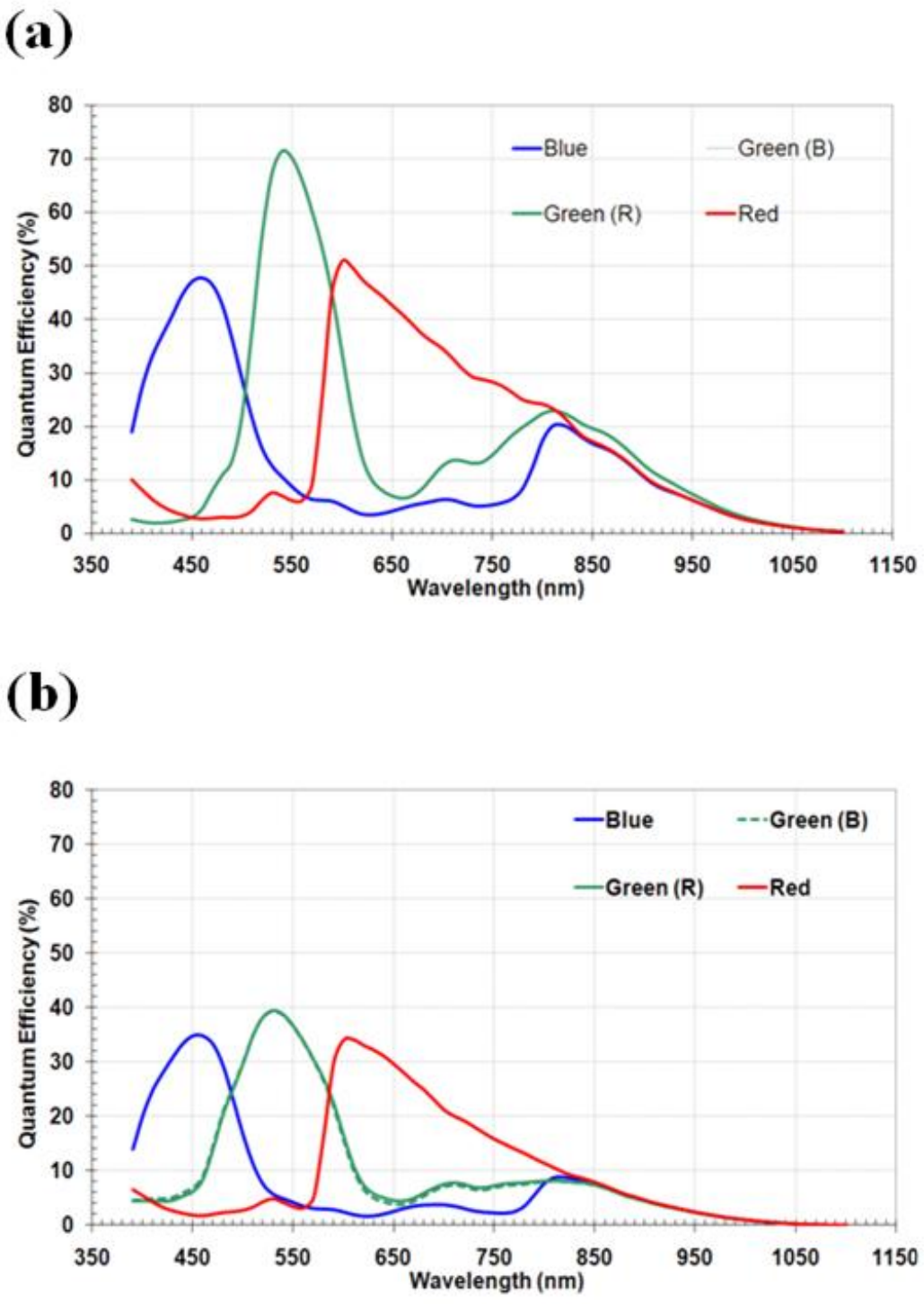


Figure 4.15: *Measured quantum efficiencies of (a) dual-bucket pixel (b) quad-bucket pixel*

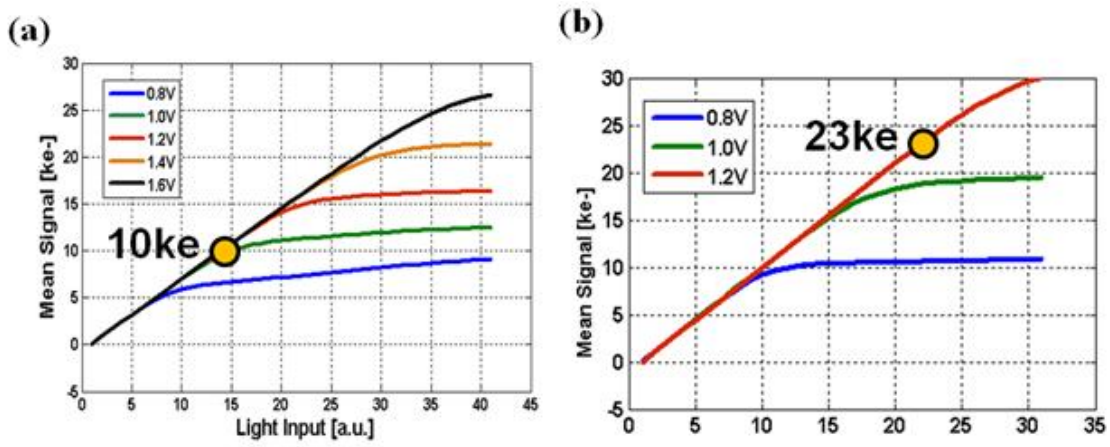


Figure 4.16: *Measured photo-response of (a) dual-bucket pixel and (b) quad-bucket pixel with respect to different SG mid-level voltages.*

the surface-generated dark current under the storage gate. Currently, we transfer and store charges along the Si/SiO_2 interface where there are surface traps that facilitate the dark current generation [Theuwissen 95]. The higher the voltage we apply to the SG, the more these traps are activated and therefore the higher the dark current.

The operating voltage is therefore a trade-off between storage capacity and dark current. One can take advantage of this relationship to adaptively bias the gate such that a low SG voltage is used under low light situation to mitigate the dark current while a high SG mid voltage is used under strong light situation to maximize the pixel capacity. To illustrate this point, Figure 4.18 shows the quad-bucket pixel's signal-to-noise ratio (SNR) measured at different signal levels with respect to different SG mid-level voltages. As shown in Figure 4.18(a), SNR drops when signal level approaches the bucket capacity. The reason is that there is always a pixel-to-pixel variation in bucket capacity due to process variations. When the signal level is getting

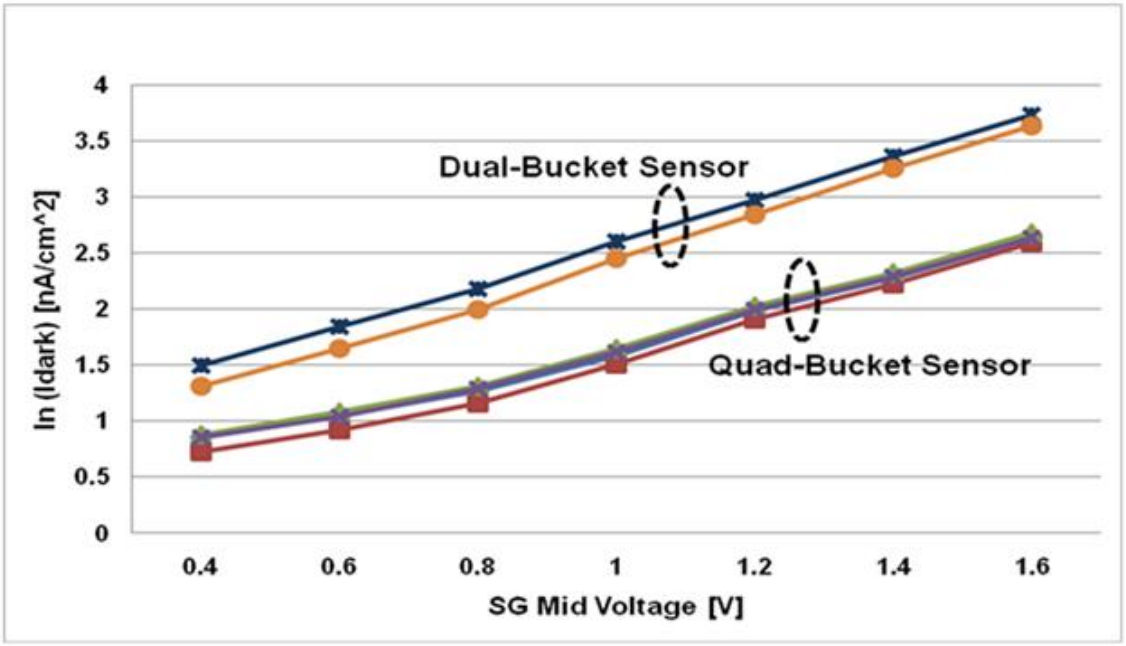
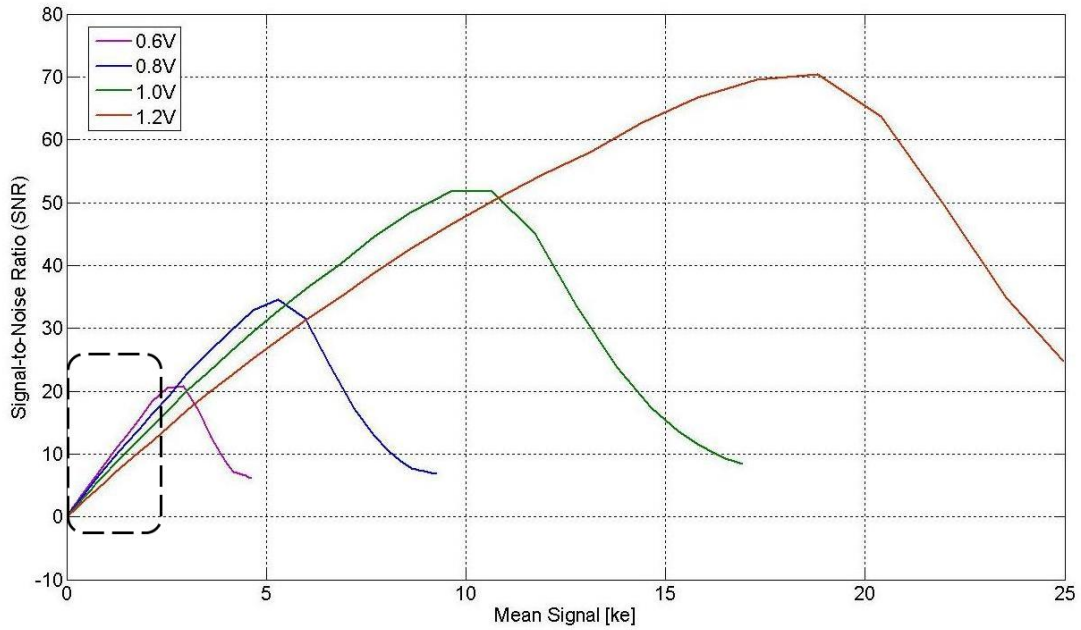
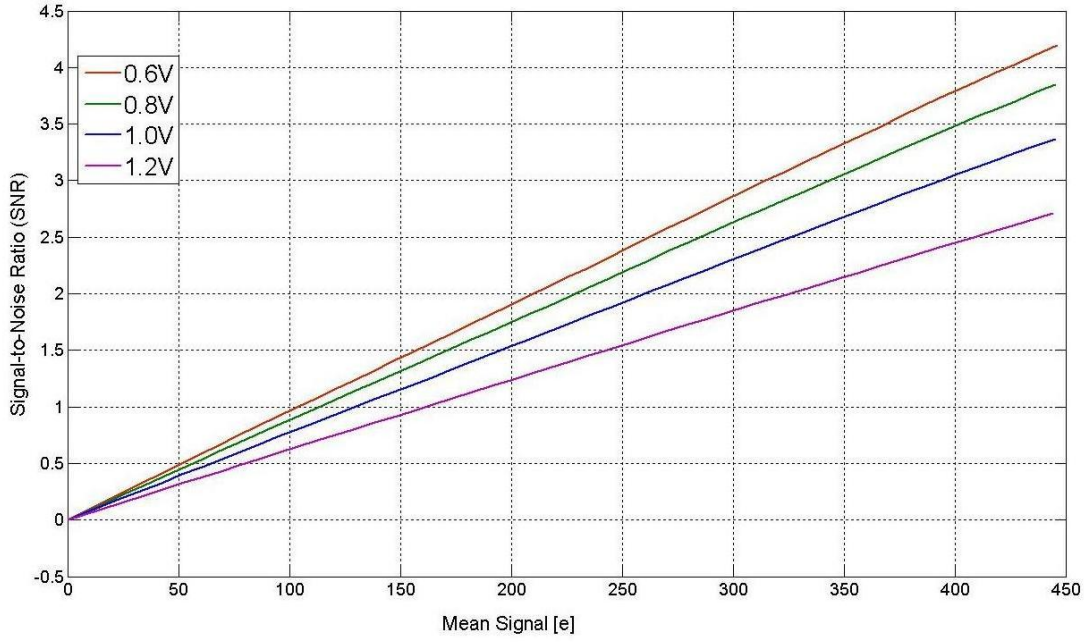


Figure 4.17: *Measured SG dark current with respect to different SG mid-level voltages at room temperature. The different curves represent dark currents of different SGs.*

closer to the bucket capacity, buckets of some pixels start to saturate while others do not. Such a difference in pixel response corresponds to image noise and therefore a drop in SNR. A larger SG mid voltage corresponds to a bigger bucket capacity and therefore a higher maximum SNR. However, a smaller SG mid voltage corresponds to smaller dark current. Therefore, as we can see in Figure 4.18(b), the SNR is higher for a smaller SG mid voltage at low light situation. Figure 4.19 shows some captured images to illustrate the capability of this adaptive biasing scheme in handling two extreme lighting conditions. As we can see, by using a smaller SG mid voltage (0.8V), the low light image is less noisy than the other that uses a larger SG mid voltage (1.2V). On the other hand, the strong light image captured using the smaller SG mid voltage has already been saturated and therefore much noisier than the other one



(a)



(b)

Figure 4.18: *Signal-to-Noise ratio at different SG mid-level voltages at room temperature. (a) Full-scale signal range (b) Zoom in of circled region in (a).*

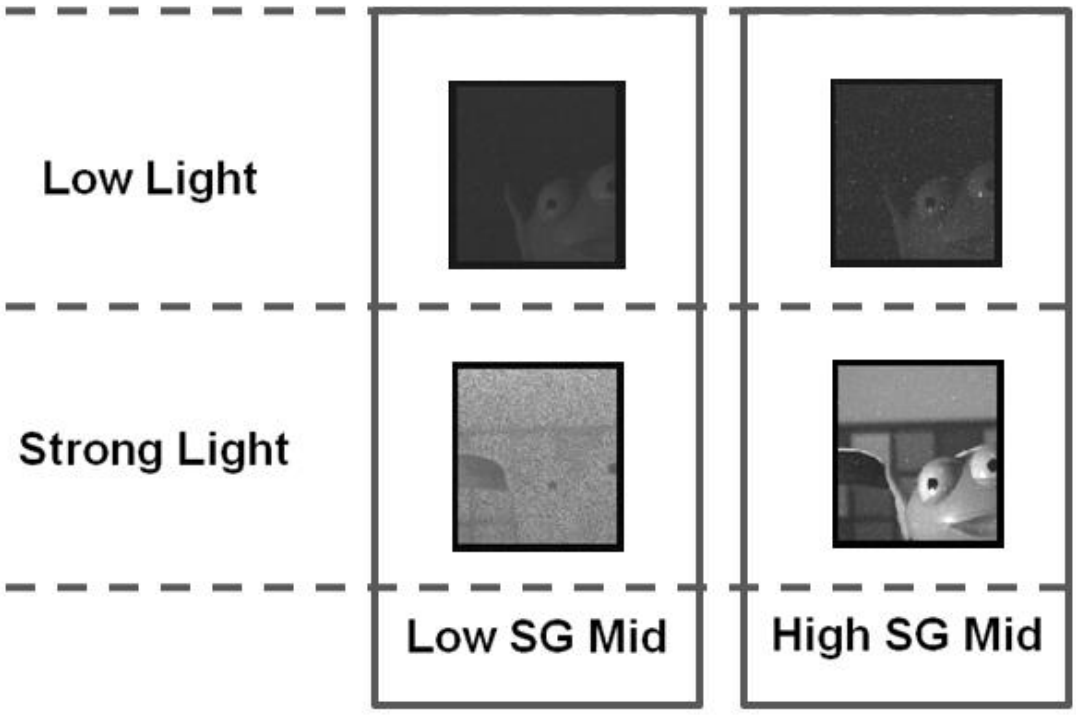


Figure 4.19: *Illustration of adaptive gate biasing scheme in handling extreme lighting conditions (Images captured by the quad-bucket sensor).*

that uses the larger SG mid voltage.

After verifying the functionalities of the sensors and seeing some of their characteristics, we will demonstrate their values by showing some computational photography results in the next chapter.

Chapter 5

Applications in Computational Photography

In this chapter, we present some computational photography applications enhanced by the multi-bucket sensors presented in the previous chapter. We will see that the sensors, through enabling time-multiplexed exposure, help eliminate many artifacts that multi-image computational photography faces when a conventional sensor is used.

5.1 Flash/No-Flash Imaging

In flash/no-flash imaging [Petschnigg 04] [Eisemann 04], two images, one with and the other without a flash, are taken to synthesize an image which preserves natural

ambient illumination while simultaneously captures dark regions of a scene. Nevertheless, this technique requires the two photographs to be taken as rapidly as possible to avoid relative motion between the camera and the scene. In [Petschnigg 04] [Eisemann 04], the experiments performed without using a tripod all require good image alignments. However, registering the flash and no-flash images are non-trivial because the lighting conditions are dramatically different and further development of more robust techniques to register such image pairs were suggested by the authors.

The dual-bucket sensor overcomes this limitation by eliminating the need to perform any image alignment. By alternating between a LED flash and no-flash and synchronizing the flash with one of the SGs as shown in Figure 5.1, one of the buckets captures a scene illuminated by the flash while the other captures the scene due to the ambient light only. The four windows separated by black lines in the images correspond to pixels with slightly different designs. Compared with a conventional sensor, our dual-bucket sensor produces two images representing the same span of time and having roughly the same motion as demonstrated in Figure 5.2. The letter S attached on the oscillating metronome needle has exactly the same blur in both flash/no flash images.

5.2 High Dynamic Range Imaging

Most high dynamic range (HDR) applications use multiple exposures [Debevec 97] [Reinhard 06], with varying exposures to capture the true dynamic range of the scene and tone map this data to synthesize a HDR image. This approach requires the

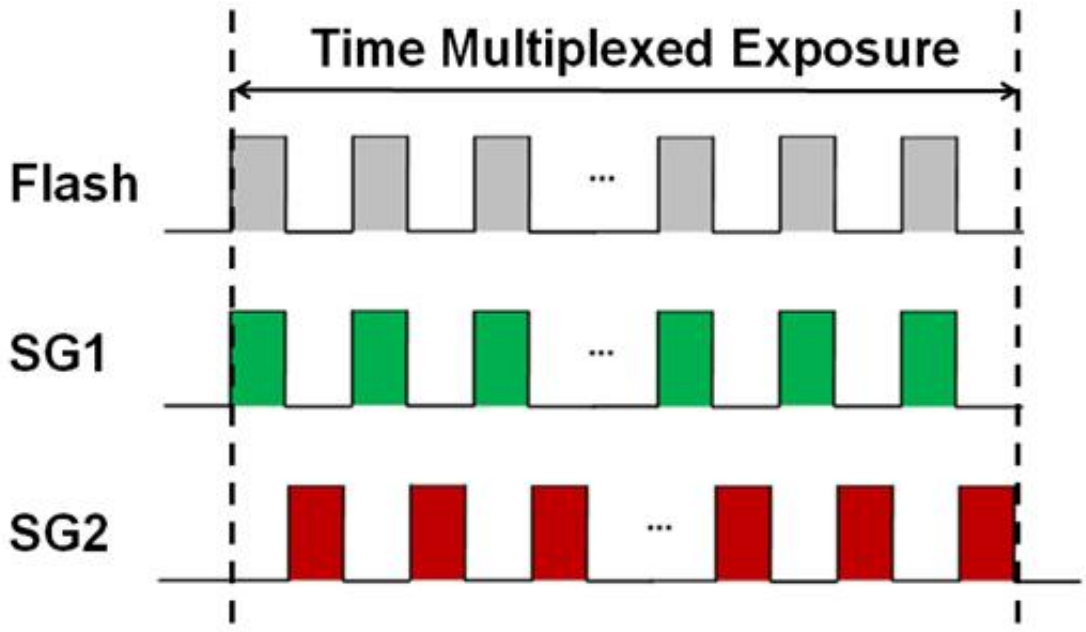


Figure 5.1: *Timing diagram for the flash/no-flash imaging.*

scene to be relatively stationary or otherwise ghosting artifacts would appear in the synthesized image [Gallo 09] [Ward 2003] [Kang 03]. The dual-bucket sensor handles this situation by fluttering the SGs so that the two buckets are recording over the same time interval, with the duty cycle of the SG clock setting the desired exposure ratio as shown in Figure 5.3. Since the two images are interleaved in time, similar motion blurs appear in the two captured images (i.e. the letter "S" attached on an oscillating metronome) as shown in Figure 5.4. The two time-interleaved images have an exposure ratio of 5:1 and there is no ghosting artifact observed in the synthesized image as shown in Figure 5.5.

By capturing four instead of two time-interleaved images, the quad-bucket sensor can span a higher dynamic range or reduce signal-to-noise ratio (SNR) dips at

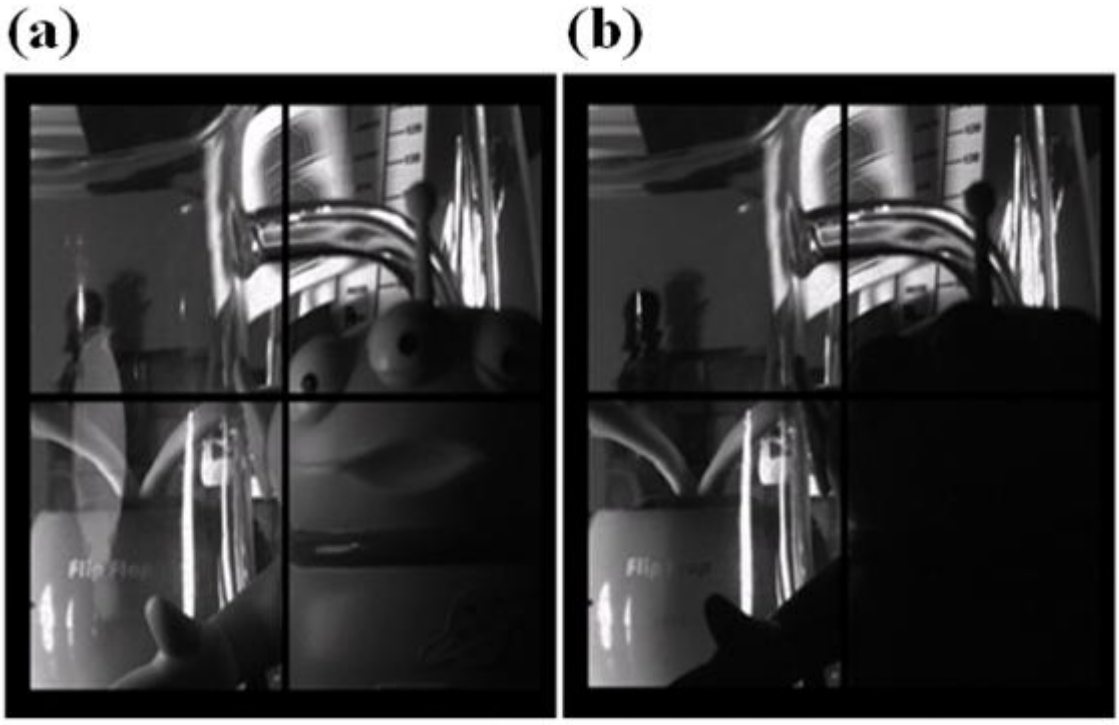


Figure 5.2: *Flash/No-Flash imaging. The scene is illuminated by a pulsing LED flash. A toy alien and an apple are placed in front of and behind a glass beaker, respectively. The letter S attached on an oscillating metronome needle placed at the back shows the same motion blur in both flash/no-flash images. (a) SG synchronized with the flash (b) SG not synchronized with the flash*

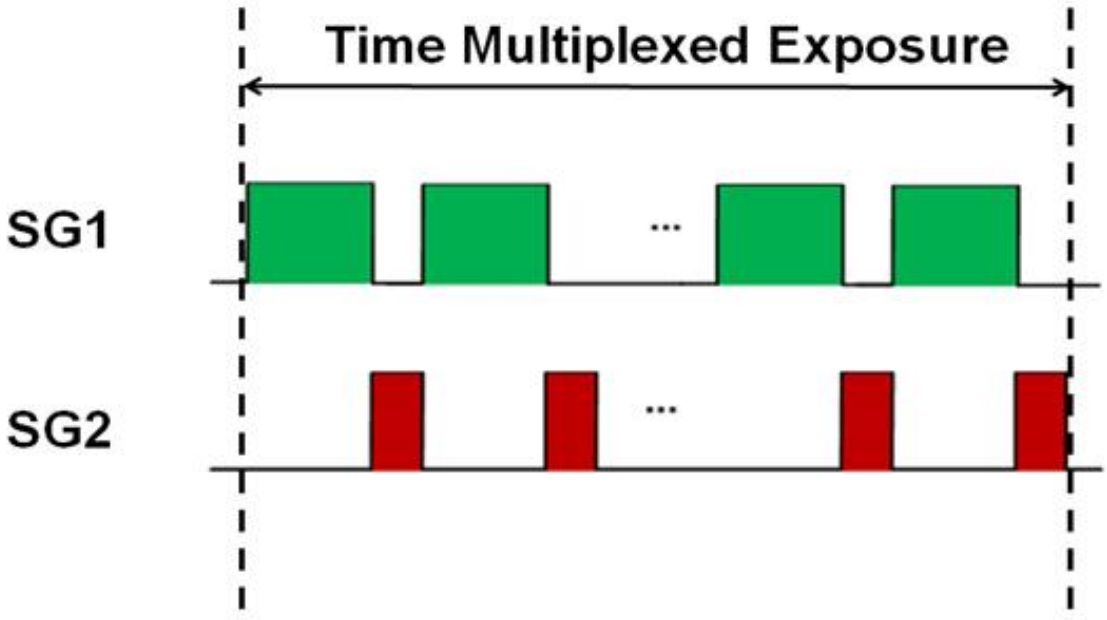


Figure 5.3: *Timing diagram for the dual-exposure HDR imaging. The duty cycles of the two SGs can be programmed to set the desired exposure ratio of the two captured images.*

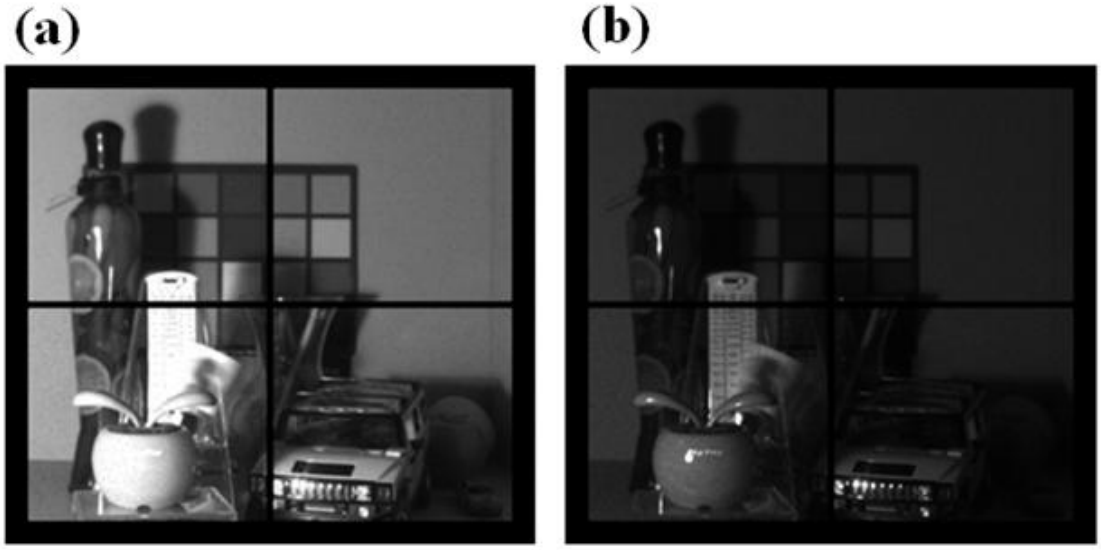


Figure 5.4: *Time-interleaved HDR imaging. The letter S attached on an oscillating metronome needle has the same blur in the two captured images. (a) Long exposure (b) Short exposure.*

Figure 5.5: *Synthesized HDR image (5:1 exposure ratio).*

transition points between exposures [Yang 99] compared with the dual-bucket sensor. Figure 5.6 shows that, during an exposure, the duty cycles of the four SGs are set to a desired ratio, e.g. 8:4:2:1, so the captured images correspond to different exposure times. As we can see in Figure 5.7, the four captured images appear as if they are captured simultaneously and therefore the synthesized HDR photography does not show ghosting nor color artifacts even without any image alignment nor complicated motion compensation algorithms. As shown in Figure 5.8, this sensor spans 85dB dynamic range with 3dB SNR dip at each transition point. A 103dB dynamic range with 6dB SNR dip can be achieved if the ratio between each of the four exposures is changed from 2 to 4.

5.2.1 Single-Exposure Time-Interleaved HDR Imaging

Although the time-interleaved HDR approach described above has shown to be effective in overcoming motion and color artifacts, an associated drawback of the approach, as described in Section 2.4, is that each captured image spans a longer absolute time. Consider the case of time-interleaved quad-exposure HDR, assume T1, T2, T3, and T4 are the desired exposure times and that T1 > T2 > T3 > T4. Further assume that each of the four exposures is partitioned into N pieces which are interlaced periodically. The captured images would then span T1+(N-1/N)(T2+T3+T4), T2+(N-1/N)(T1+T3+T4), T3+(N-1/N)(T1+T2+T4), and T4+(N-1/N)(T1+T2+T3). Especially when N is large or the exposure ratio is small, the increases in time spans of the images make them more susceptible to motion blur. In the followings, we propose

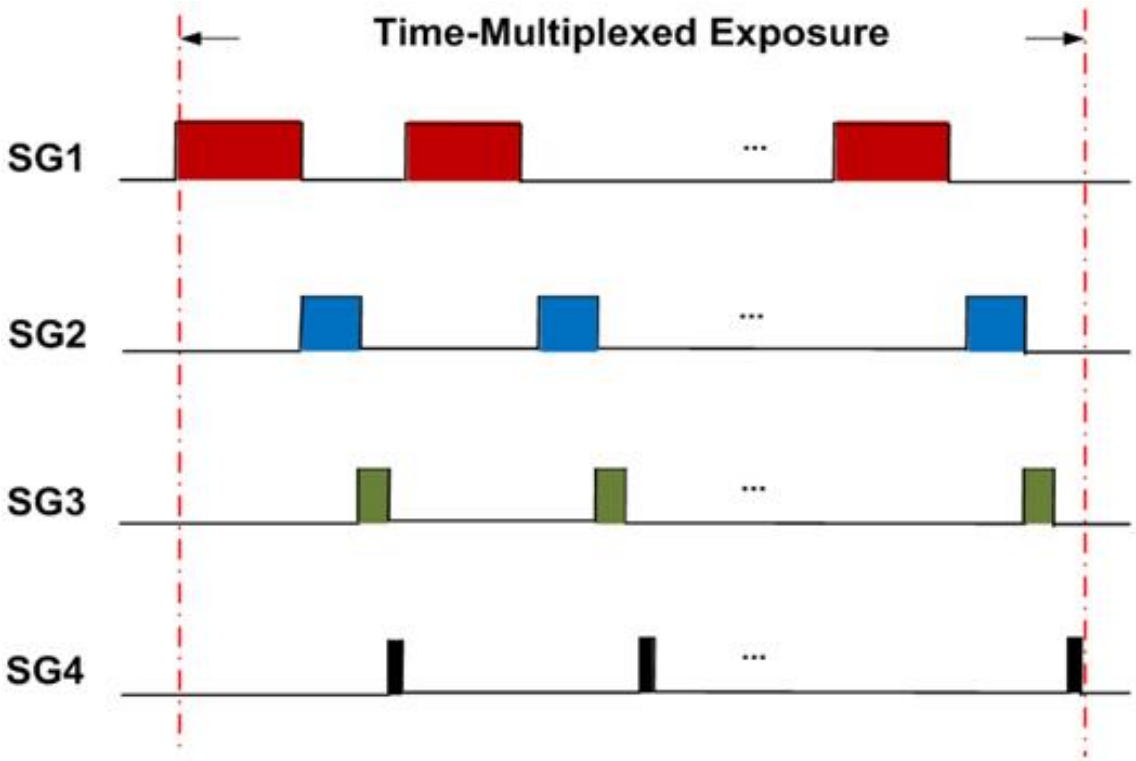


Figure 5.6: *Timing diagram for the quad-exposure HDR imaging.*

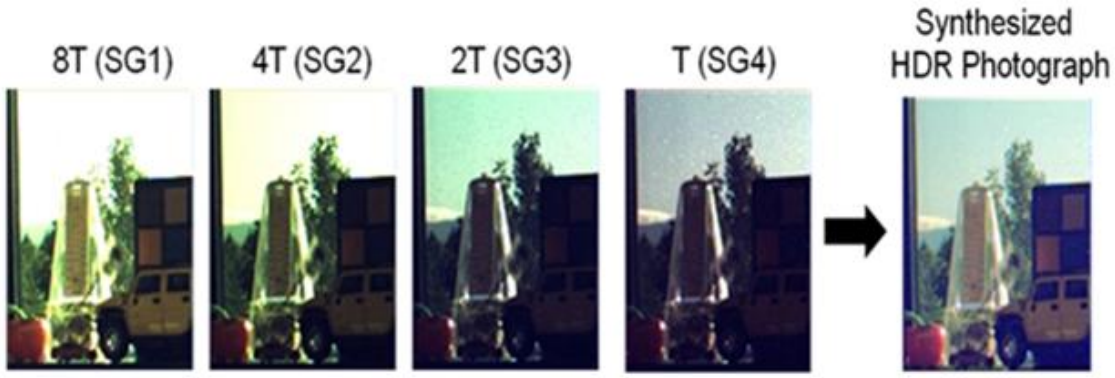


Figure 5.7: *Time-interleaved quad-exposure HDR imaging. The metronome needle and trees are moving in the scene. The sensor captures these four images virtually simultaneously thus eliminating the need to align the images when synthesizing the HDR photograph.*

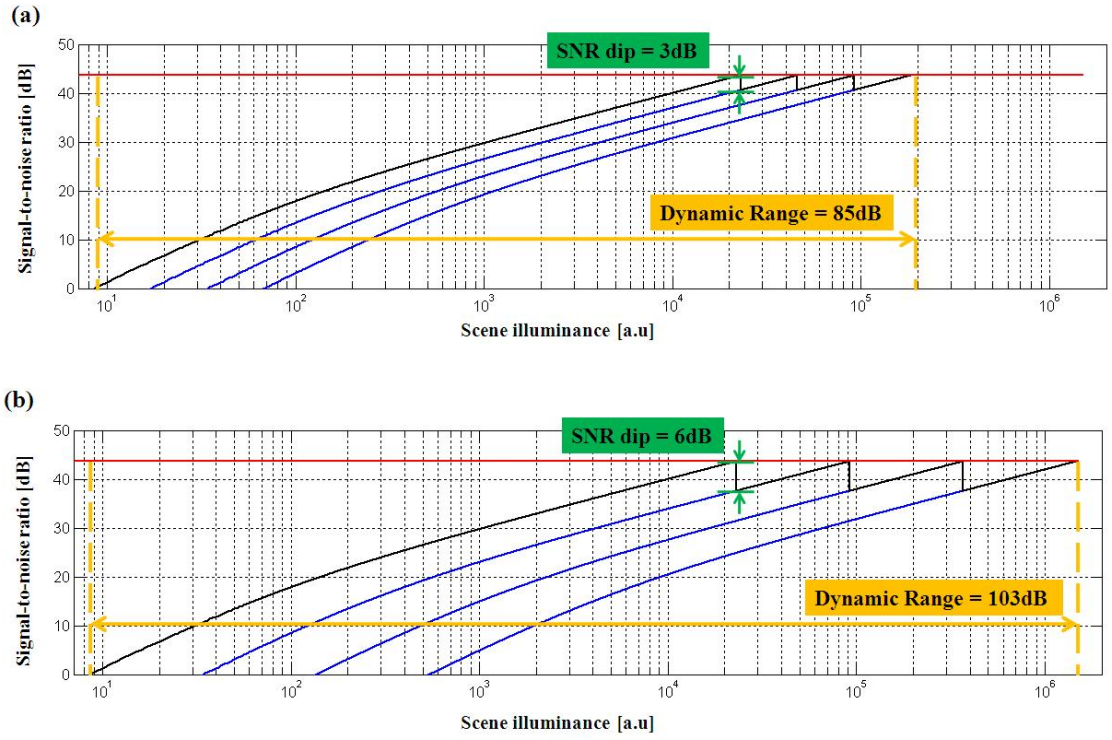


Figure 5.8: *Signal-to-noise ratio (SNR) of images associated with quad-exposure HDR imaging. In (a), the ratio between each of the four exposures is 2 while that in (b) is 4. The maximum achievable SNR (red line) is determined by the square root of the bucket capacity, i.e. 43.6dB. The blue lines show the SNR of each component image used to synthesize the final HDR image. The black lines show the corresponding SNRs of the final HDR images. In (a), the dynamic range is 85dB with 3dB dip at each transition point. In (b), the dynamic range is increased to 103dB with a tradeoff of an increased 6dB SNR dip at each transition point.*

a new time-interleaved HDR approach that removes this drawback completely. It should be noted that, even though this new approach has not been verified experimentally at the time of writing this thesis, both the dual and quad-bucket sensors developed in this thesis are capable of implementing the proposed approach.

Figure 5.9 illustrates how the new time-interleaved HDR works by considering the same example of capturing four frames with exposure times 8T, 4T, 2T, and T, where T is an arbitrary unit of time. Instead of setting the duty cycles of the SGs to be 8:4:2:1 like in the conventional time-interleaved HDR, in this new approach SG1:SG2:SG3:SG4 are instead set to be 4:2:1:1. As shown in the figure, the image corresponding to exposure time T is readout directly from the bucket SG4 while that corresponding to 2T is obtained by summing the signals captured by buckets SG4 and SG3. Similarly, the image corresponding to 4T is the sum of the signals captured by buckets SG4, SG3 and SG2. Finally, the image corresponding to 8T is the sum of the signals captured by all four buckets. As we can see from the figure, the total capture time is shortened from 15T to 8T and the image that corresponds to the longest exposure, i.e. the one captured by SG1, now spans an absolute time of 8T instead of 8T+(N/N-1)(7T). In fact, since the longest exposure itself requires 8T, this new approach takes the theoretical minimum amount of time to finish capturing the multiple images. As multiple time-interleaved images can be obtained using simply one single exposure time, we call this new approach ***single-exposure time-interleaved HDR imaging***. Besides achieving the minimum total capture time and the optimal absolute time span of the longest exposure, this new approach has an

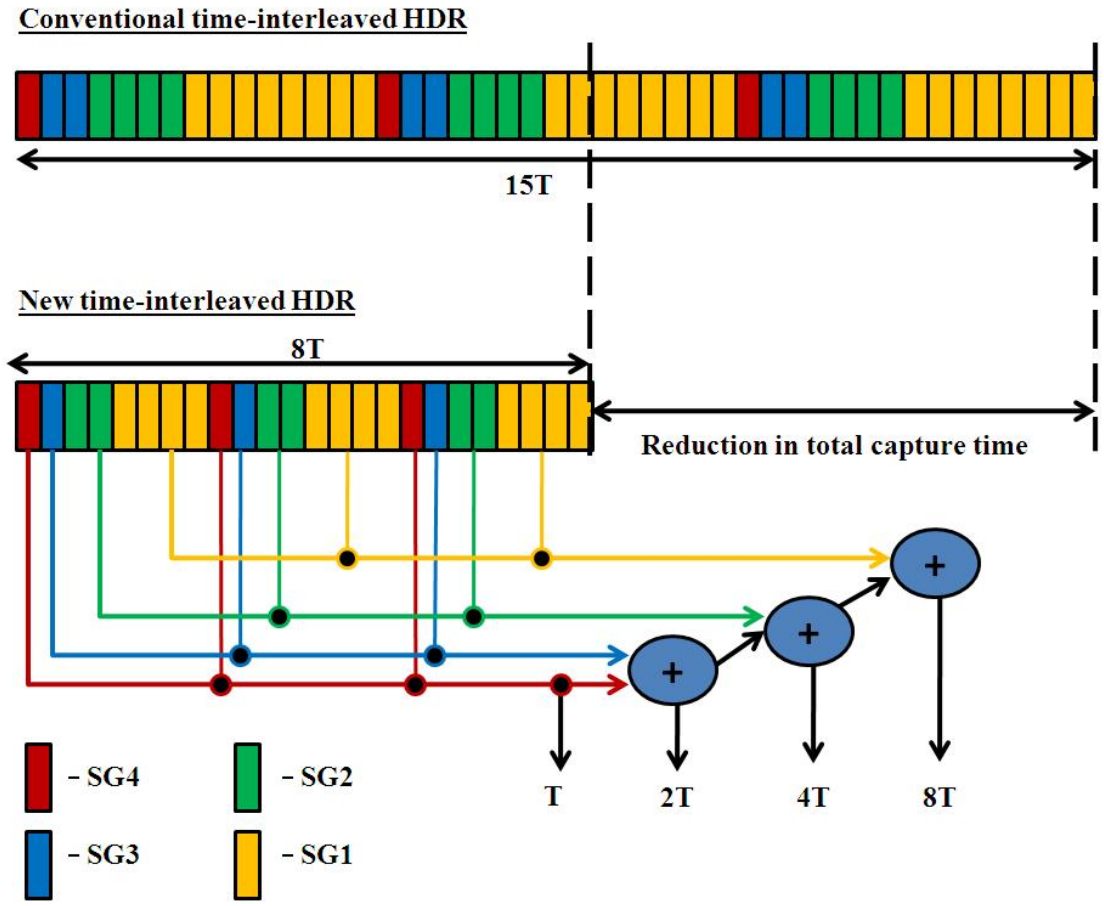


Figure 5.9: *Single-exposure time-interleaved HDR Imaging. In this approach, the image corresponding to exposure time T is readout directly from the bucket SG4 while that corresponding to 2T is the sum of the signals captured by buckets SG4 and SG3. Similarly, the image corresponding to 4T is the sum of the signals captured by buckets SG4, SG3 and SG2. Finally, the image corresponding to 8T is the sum of the signals captured by all four buckets. Compared with the conventional time-interleaved imaging, this new approach achieves the theoretical minimum total capture time, has the optimal absolute time span of the longest exposure, and brings the centroids of images even closer.*

additional benefit that the centroids of the images are closer than those in conventional time-interleaved HDR as we can clearly see from the figure.

The signals from the buckets can be summed after they are readout but it would be more optimal to sum them at the floating diffusion of the pixel from the signal-to-noise (SNR) perspective. Similar to accumulating signals at the buckets, adding signals at the floating diffusion in the charge domain is in principle a noiseless process. Also, as discussed in Section 2.6.1, summing signals before they are readout reduces the penalty from the finite read noise and therefore improving SNR of the final image compared with if the signals are added after they are readout.

5.3 Color Imaging using Active Illumination

Another application of our quad-bucket sensor is color imaging using active illumination without a color filter array (CFA). One of the approaches to achieve color imaging is illuminating a scene sequentially using three light sources Red (R), Green (G), and Blue (B) and taking three corresponding pictures which are combined to form a final color image [Ohta 07]. Given excellent color fidelity due to the lack of cross-talk and improved light sensitivity by eliminating the CFA, this approach is attractive in light-limited applications such as capsule endoscopy that requires good color reproduction and high light efficiency [Ohta 07]. However, the conventional way of doing this usually suffers from severe color artifacts due to motion between the three exposures [Xiao 01]. Figure 5.10 shows how the quad-bucket pixel overcomes this artifact. Three time-interleaved RGB light sources are used to illuminate the

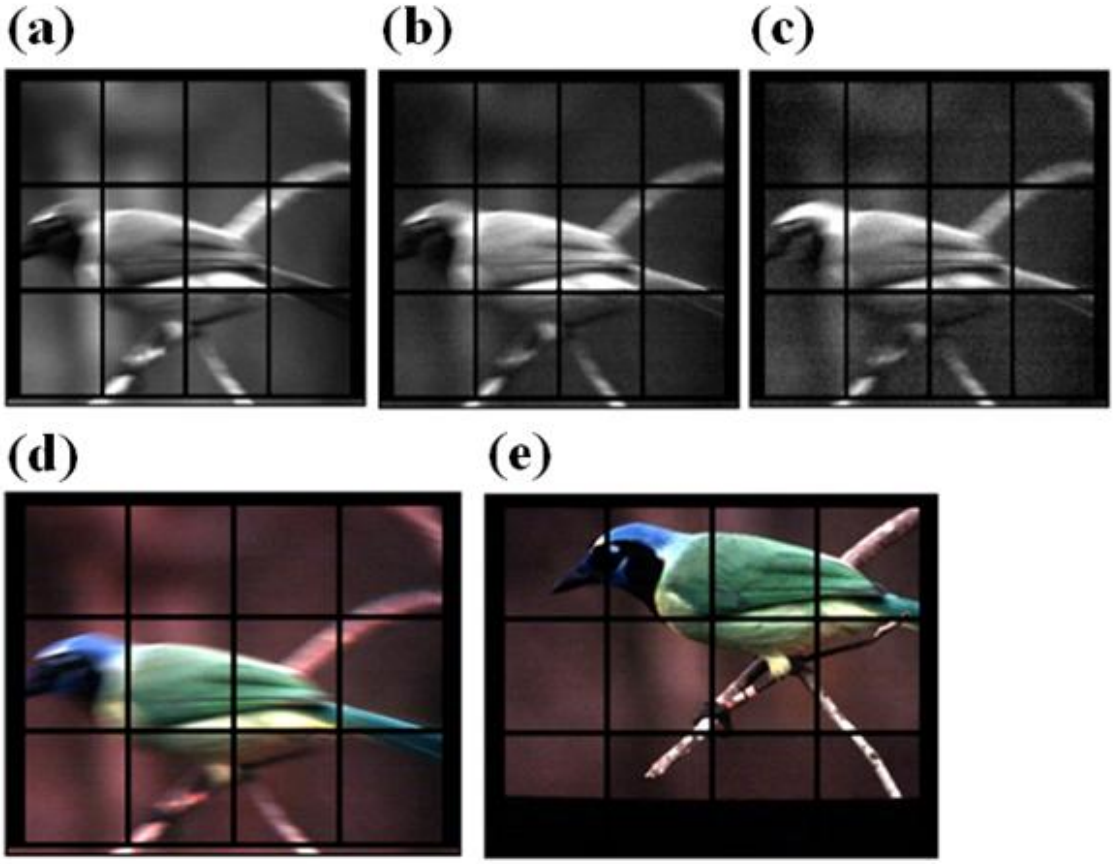


Figure 5.10: *CFA-less color imaging. (a) Red (b) Green (c) Blue images captured by SG1, SG2, and SG3 synchronized with R, G, and B light sources, respectively. (d) Synthesized color image. Color artifact due to motion is not observed. (e) Static image.*

scene while three of the four SGs are synchronized with the light sources. Three time-interleaved RGB images are obtained and are combined to form a final color picture. Color artifacts commonly observed when a conventional sensor is used are removed due to the time-interleaved nature of the captured images. Optionally, the remaining SG can be used to capture an image corresponding to a fourth lighting condition such as the ambient light or a scene being illuminated by a fluorescent LED in the case of capsule endoscopy applications.

5.4 Multi-Flash Imaging

Multi-flash imaging has been proposed for depth edge detection and non-photorealistic rendering [Raskar 04]. Several light sources are used to illuminate an object and images, each corresponds to when one of the light sources is on, are captured. Since the light sources are placed at different positions relative to a camera, they cast different shadows of the object onto a background. By computationally combing those images, a final shadow-free image can be synthesized which can be used to recover the depth edge of the object. In [Raskar 04], the authors explicitly stated that not being able to simultaneously capture the scenes with moving objects is a limitation of this imaging approach. Although motion compensation techniques can be used to correct artifacts introduced to a certain degree, the assumptions need to be made on the scene are fairly restrictive. For example, when there are thin objects or objects with high frequency texture, large motions between successive frames would result in spurious edges even when their proposed algorithm is used [Raskar 04]. Therefore, there remains a challenge for this imaging technique to be used in real scene.

Figure 5.11 shows the experimental setup used to demonstrate the quad-bucket sensors capability to eliminate this limitation through performing time-interleaved multi-flash imaging. Four groups of white LEDs, located at the top, bottom, left, and right of the sensor, are turned on sequentially and repeatedly during an exposure with each group of LEDs being synchronized with one of the buckets as shown in Figure 5.12. To a conventional sensor and human eyes, the four groups of LEDs switch so rapidly that they appear to turn on and off simultaneously (Figure 5.13).

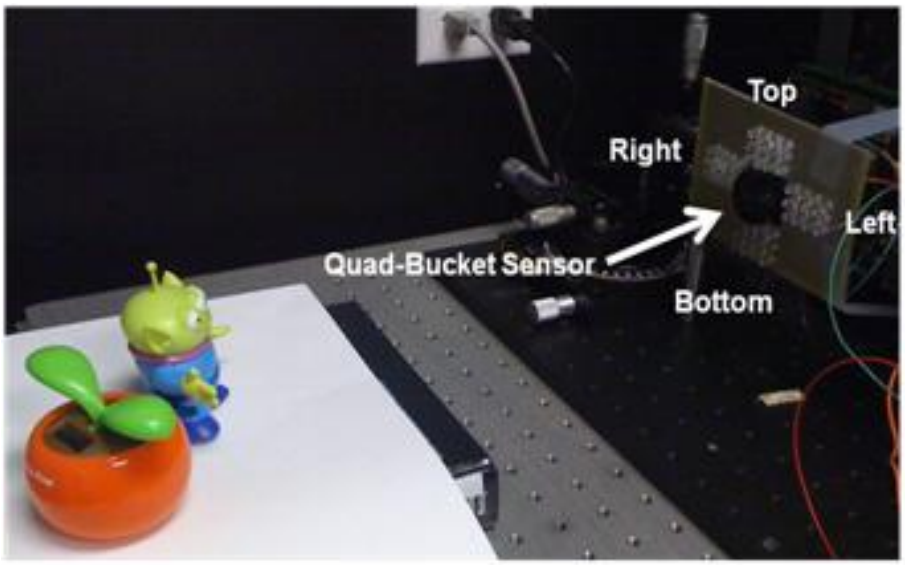


Figure 5.11: *Experimental setup of time-interleaved multi-flash imaging. Four groups of white LEDs, located at the top, bottom, left, and right of the sensor, are turned on sequentially and repeatedly during an exposure with each group of LEDs being synchronized with one of the buckets in the quad-bucket sensor.*

To a quad-bucket sensor with its buckets synchronized with the LEDs, however, four time-interleaved images with different shadows are captured as shown in Figure 5.14. A final shadow-free image computed using the four captured images is shown in Figure 5.15. In contrast to a conventional sensor, which captures the four images sequentially, the quad-bucket sensor time-interleaves the captures and makes the imaging approach more robust to motion in the scene.

5.5 Flash Matting

Flash matting utilizes the fact that a flash brightens foreground objects more than the distant background to extract mattes from a pair of flash/no-flash images [Sun 06].

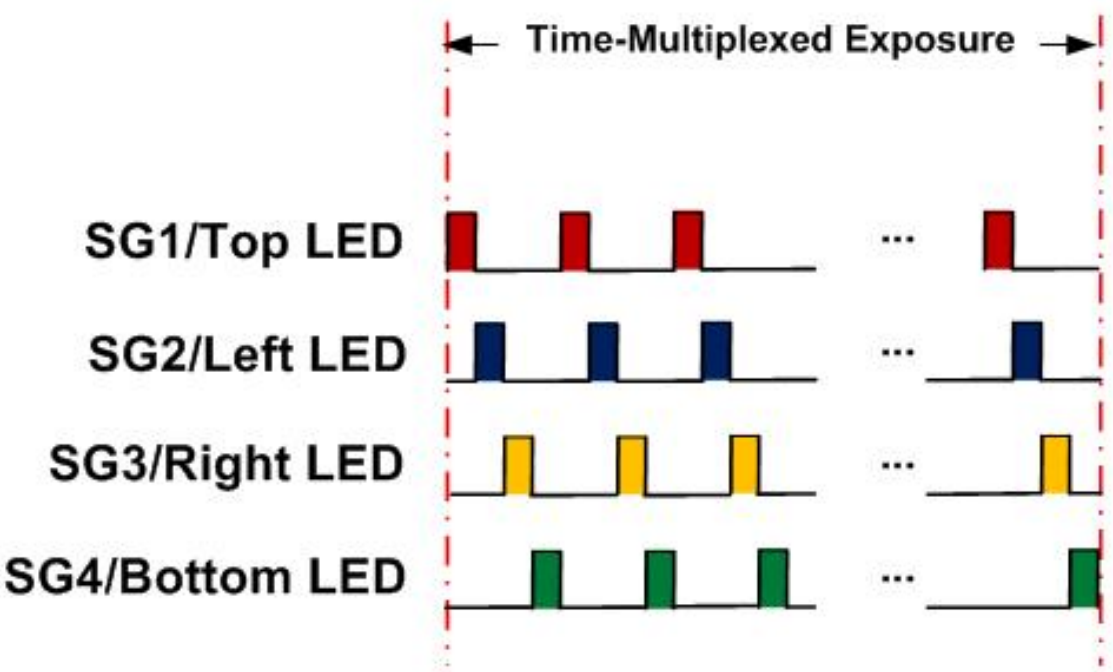


Figure 5.12: *Timing diagram of the time-interleaved multi-flash imaging. Each group of LED is synchronized with one of the buckets (or SG) in the quad-bucket sensor.*

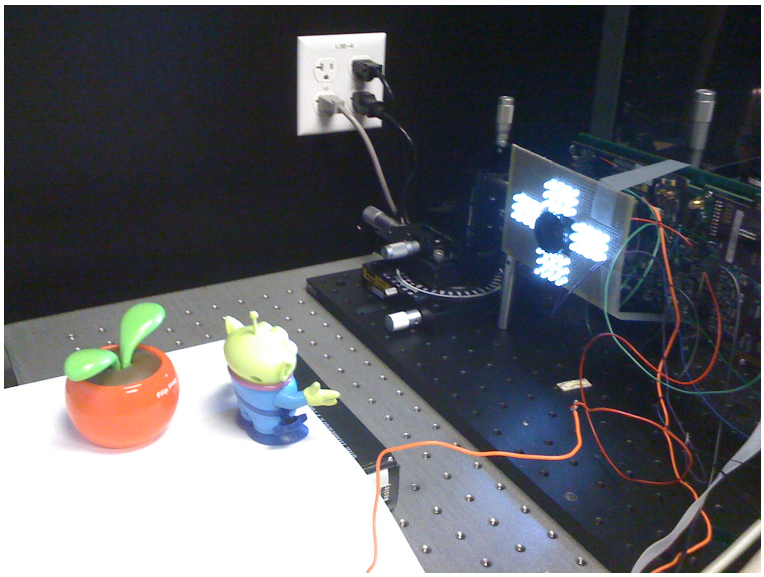


Figure 5.13: *An image of flashing LEDs taken by a conventional sensor. Since the four groups of LEDs switch on and off very rapidly, they appear to turn on and off simultaneously to a conventional sensor that was used to take this picture.*

This approach works nicely even when the foreground and the background are similar or the background is complex. However, one of the assumptions for this technique to work correctly is that the input image pair needs to be pixel aligned. Although the technique was later improved by combining flash, motion, and color cues in a Markov Random Field (MRF) framework [Sun 07], still only moderate amounts of camera or subject motion can be handled. Therefore, the approach is currently restricted to be used in a practically static scene.

Both of the dual and quad-bucket sensors, when combined with a flash, can be used to perform flash matting for a dynamic scene. One of the buckets is used to record an image corresponding to when a flash is illuminating foreground objects while other buckets capture the scene when the flash goes off. Again, because our

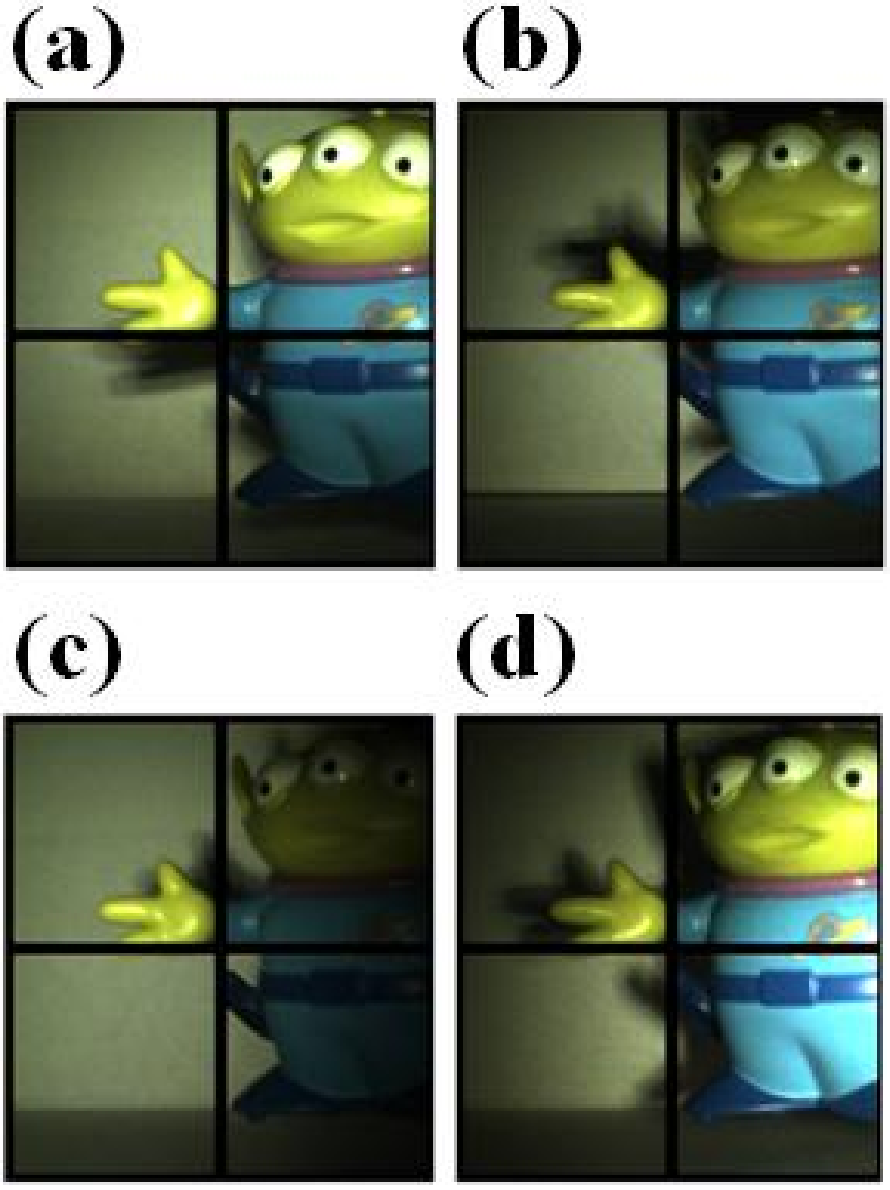


Figure 5.14: *Multi-Flash imaging. Images captured when (a) Top (b) Bottom (c) Left (d) Right LED groups are illuminating the scene.*

Figure 5.15: *Synthesized shadow-free Image.*

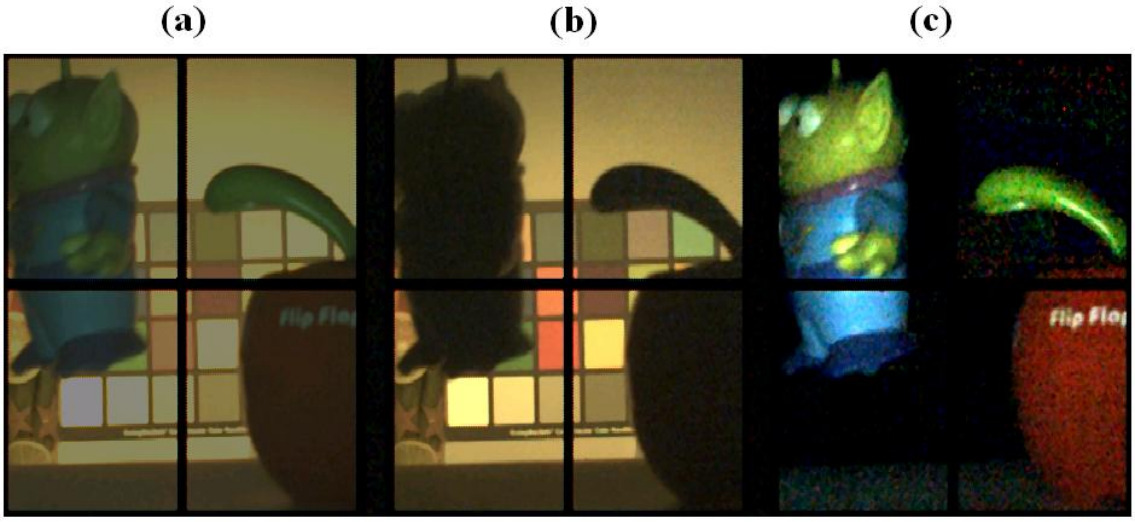


Figure 5.16: *Flash Matting. In this scene, the toy alien is moving while the leaves of the toy apple are waving up and down. (a) Background + foreground image when the flash is on (b) Background only when the flash is off (c) Extracted foreground objects.*

sensors allow interleaving the captures, motion artifacts or other changes in the scene are effectively suppressed. By simple arithmetic operations on the captured images, we can compute a final image which shows only the foreground objects as shown in Figure 5.16.

From the various applications described, we can see that the multi-bucket sensors, through enabling time-multiplexed exposure, eliminates the need for image alignment when combining images in multi-image computational photography and therefore avoids artifacts that would potentially arise when a conventional sensor is used.

Chapter 6

Conclusion

Looking back in the history of photography, we see several important changes. From the early camera obscura to the film-based camera and more recently the digital camera, remarkable improvements have been made in image quality, cost, and popularity. However, the underlying image formation process remains essentially the same - a medium such as a film or an image sensor records an image projected onto it. Computational photography promises to revolutionize the concept of photography by altering the image formation process such that intermediate images captured by the recording device are used to compute the final image.

Multi-exposure discussed in this thesis is one of the most commonly used techniques in computational photography. Despite demonstrated success, today this technique can only be used in limited situations. The reason is that changes in the scene during the multiple captures, particularly motion, cause undesired differences in the captured images. Being unpredictable, the differences usually cause the subsequent

reconstruction algorithms to fail resulting in artifacts in the final computed image. This situation is especially true when the algorithms are those that do not involve human intervention. To extend the application domains of this computational photography technique, this research has taken an alternative approach to fundamentally change the way we capture multiple images.

In this thesis, we have described time-multiplexed exposure within which multiple images can be captured in rapid succession or in a tightly time-interleaved manner. Compared with the conventional way of capturing images sequentially limited by the frame rate of the sensor used, time-multiplexed exposure eases the subsequent error-prone image alignment algorithms or eliminates them altogether. Undesired changes in the scene that currently plague multi-image computational photography are therefore effectively removed. In this imaging approach, memory is indispensible because intermediate sub-images need to be stored. In principle, one can put the memories external to a faster conventional image sensor or implement analog memories inside each pixel. The former approach is infeasible using existing technologies due to its demanding requirements in sensor frame rate, data rate, and power consumption. The later approach eliminates these demands but requires a multi-bucket pixel that has two or more analog memories.

Designing a multi-bucket pixel is challenging because additional pixel areas are required to implement the memories. It is especially true if the pixel needs to simultaneously be small, preserve correlated double sampling, have low dark current, and allow shared pixel architecture. In this thesis, we have adapted the old idea of

virtual phase CCD into a modern CMOS technology to design a new pixel structure that forms the basis of our multi-bucket pixels. The pixel structure possesses all the desirable features listed above, except low dark current. Based on this proposed pixel structure, we have designed and fabricated two prototype multi-bucket CMOS image sensors with dual and quad-bucket pixels, both of which have the smallest sizes among similar pixels reported in the literature. Various design tradeoffs are explored and characterization results agree with predictions. Currently, the biggest concern of this pixel is its dark current performance. Although it can be improved by optimizing the process conditions, the pixel in its current form would fundamentally have higher dark current when compared with a state-of-the-art 4-transisor pixel, mostly because of the surface-generated dark current at the storage node.

Finally, using the multi-bucket sensors, we implemented various computational photography applications to demonstrate the value of time-multiplexed exposure in avoiding artifacts that plague those applications by completely eliminating the need for image alignments.

6.1 Final Thoughts and Future Work

Computational photography has its potential to cause a paradigm shift in the way people do photography. So far, due to its roots in computer graphics, computer vision and applied optics, almost all of the researches in computational photography focus on modifying the optics or improving the reconstruction algorithms. Despite the fact that an image sensor has always been the industry's focus in improving camera

performance, there is only very limited research in developing new image sensors for computational photography. The work described in this thesis is one of the first attempts to tap into this less-explored area.

Time-multiplexed exposure described in this thesis has shown to be effective in completely avoiding reconstruction artifacts that can potentially arise in multi-image computational photography. However, except for the proposed single-exposure time-interleaved high dynamic range imaging described in Section 5.2.1, time-multiplexed exposure comes with the cost that the time span of each image becomes longer and therefore more susceptible to motion blur. While motion blur due to camera shake can potentially be avoided by using image stabilization hardware or a tripod, object motion blur cannot be tackled this way. In some cases, such a motion blur is acceptable or preferable, for instance in intentionally creating artistic effects. In some other cases, however, the blur can be a nuisance. If it is the case, one may prefer to simply use the multi-bucket sensors to capture the multiple images very rapidly without interleaving and subsequently align them with a reduced image alignment complexity. However, the numbers of images that can be captured are limited by the number of buckets per pixel. Additionally, if motion blur exists within an image, for example when there is a very fast moving object in the scene, aligning the images may still be difficult. Therefore, an interesting future research would be to implement the concept of coded-exposure or flutter shutter [Raskar 06b] into the multi-bucket sensors by carefully coding the pattern of charge transfer to the buckets. Since high

frequency components are preserved, the captured images can in principle be deblurred. However, reconstruction artifacts may occur after the deblurring process so one needs to carefully study and experiment the pros and cons of this approach. In any case, the multi-bucket sensors developed in this research provide experimental platforms for this study.

Of course, another option would be to build a very fast sensor together with a supporting system. In this case, we can readout many images without blur by exposing each of them for a very short amount of time. As explained in the thesis, the frame and data rates required are very high and one also needs to be concerned about the resulting power consumption. From the analysis given in this thesis, it is unlikely that such a system will appear in the near future if we simply continue to incrementally improve the sensor technologies. A rethinking of the current sensor architecture and a breakthrough in power needs to happen for this approach to become practical.

In terms of the multi-bucket pixel technology, compared with the state-of-the-art pixel size (e.g. 1.4μm and 1.1μm), the multi-bucket pixels described in this research are too large to be used in applications such as smartphones. By going to a more advanced process node, pixel sizes can be made smaller. However, as discussed in the thesis, pixel transistors do not shrink with the pixel size because of some fundamental performance limitations such as RTS noise. Besides that, the width of the clocked barrier and that of the the virtual barrier pose fundamentally limitations on the size of the multi-bucket pixel because the pinned photodiode and clocked well need to be

isolated, so does the clocked well and the floating diffusion. An interesting approach to achieve a very small, multi-bucket pixel is through the use of quantum or organic films as photo-detectors. In this type of pixel, the quantum or organic film sits at the top of the pixel while the whole silicon area underneath can be used to implement the buckets and pixel transistors. In this situation, the tradeoff between photodiode and bucket sizes is completely eliminated. Another research area is reducing the dark current generated at the bucket. As we have discussed, our multi-bucket pixels have higher dark currents than state-of-the-art rolling shutter pixel. Optimizing the process would significantly reduce the dark current. However, compared with a pinned photodiode, dark current will always be higher because of the surface-generated dark current at the bucket. In addition to a new process development which requires a strong collaboration with the foundry, an even more interesting research area would be to see how we can live with the dark current by computationally correcting their effects. Currently, the effect of dark current can be removed through techniques such as dark frame subtraction, however it is sometimes not possible to perform this for example when the sensor is being used in a phone which does not have a mechanical shutter. If we know better the dark current characteristics, perhaps we can intelligently correct them based on this prior which in principle can be measured beforehand.

Finally, for computational photography researchers, the multi-bucket sensors are hardware with new functionality. It is our hope that this new piece of hardware would enable more applications and stimulate development of new software algorithms, as

well as trigger more research in developing new image sensors to enable computational photography.

Bibliography

[Agranov 09] Agranov, G., et al., "Pixel continues to shrink . . . Pixel Development for Novel CMOS Image Sensors," 2009 International Image Sensors Workshop (IISW), June 2009.

[Agranov 11] Agranov, G., et al., "Pixel continues to shrink.Small pixels for novel CMOS image sensors," Proc. Intl Image Sensor Workshop, pp. 1-4, June 2011.

[Bogaerts 11] Bogaerts J., et al., "High speed 36 Gbps 12Mpixel global pipelined shutter CMOS image sensor with CDS," Proc. Intl Image Sensor Workshop, pp. 335-338, June 2011.

[Debevec 97] Debevec, P. E., and Malik, J., "Recovering high dynamic range radiance maps from photographs," In Proc. ACM SIGGRAPH 97, T. Whitted, Ed., 369378, 1997.

[Eden 06] Eden, A., Uyttendaele, M., and Szeliski, R., "Seamless image stitching

of scenes with large motions and exposure differences," In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 24982505, New York, NY, 2006.

[Eisemann 04] Eisemann, E. and Durand, F., "Flash Photography Enhancement via Intrinsic Relighting," ACM Transactions on Graphics, 23, 3, 670-675, 2004.

[Ershov 11] Ershov, M., Cadjan, M., Feinberg, Y., Li, X., Wan, G., Agranov, G., "Accurate capacitance and RC extraction software tool for pixel, sensor, and precision analog designs," Int. Image Sensor Workshop, Hokkaido, Japan, June. 2011.

[Findlater 04] Findlater, K., Vaillant, J.M., Baxter, D.J., Augier, C., Herault, D., Henderson, R.K., Hurwitz, J.E.D., Grant, L.A. and Volle, J-M., "Source Follower Noise Limitations in CMOS Active Pixel Sensors," Detectors and Associated Signal Processing. Proceedings of the SPIE, Volume 5251, pp. 187-195, 2004.

[Furuta 06] Furuta, M., Inoue, T., Nishikawa, Y., and Kawahito, S., "A 3500fps high-speed CMOS image sensor with 12b column-parallel cyclic A/D converters," Symp. On VLSI Circuits Dig. Tech. Papers, pp. x-y, June 2006.

[Gallo 09] Gallo, O., Gelfand, N., Chen, W., Tico, M., and Pulli, K., "Artifact-free high dynamic range imaging," IEEE International Conference on Computational

Photography, April 2009.

[Grünschloß 08] Grünschloß, L., "Motion Blur," Thesis, Ulm University, 2008.

[Hayes 08] Hayes, B. Computational photography. American Scientist, 96:9499.2008.

[Holst 07] Holst G., Lomheim, T., CMOS/CCD Sensors and Camera Systems, SPIE, 2007.

[Hsu 04] Hsu, T. H., et. al., "Light Guide for Pixel Crosstalk Improvement in Deep Submicron CMOS Image Sensor," IEEE Electron Device Letters, vol. 25, No. 1, pp. 22-24, Jan. 2004.

[Hynecek 79] Hynecek, J., "Virtual phase CCD technology." in lEDM Tech. Dig., Dec. 1979, pp. 611-614.

[Janesick 07] Janesick, J.R., Photon Transfer. SPIE Press, 2007.

[Kang 03] Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R., "High dynamic range video," ACM Transactions on Graphics (Proc. SIGGRAPH), 22(3):319325, 2003.

[Kawahito 07] Kawahito, S., Halin, I., Ushinaga, T., Sawada, T., Homma, M., Maeda, Y., "A CMOS time-of-flight range image sensor with gates-on-field-oxide structure," IEEE Sens. J., 7, 12, 2007.

[Kawahito 10] Kawahito, S., "Column readout circuit design for high-speed low noise imaging," ISSCC Forum, 2010. [Kim 10] Kim, S.-J., et al., "A three-dimensional time-of-flight CMOS image sensor with pinned-photodiode pixel structure," IEEE Electron Device Lett., vol. 31, no. 11, pp. 1272-1274, Nov. 2010.

[Krymski 03a] Krymski, A. I. and Tu, N., "A 9-V/lux-s 5000-frames/s 512x512 CMOS sensor," IEEE Trans. Electron Devices, Vol. 50, No. 1, pp. 136-143, 2003.

[Krymski 03b] Krymski, A., Khaliullin, N., Rhodes, H., "A 2 e- noise 1.3megapixel CMOS sensor," IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, 2003.

[Levoy 10] Levoy, M., "Experimental Platforms for Computational Photography," IEEE Computer Graphics and Applications, 30(5), 2010.

[Leyris 06] Leyris, C., Martinez, F., Valenza, M., Hoffmann, A., Vildeuil, J.C., and Roy, F., "Impact of Random Telegraph Signal in CMOS Image Sensors for Low-Light Levels," European Solid-State Circuits Conference (ESSCIRC), pp. 376 379, Sept.

2006.

[Lotto 11] Lotto, C., Seitz, P., Baechler, T., "A Sub-Electron Readout Noise CMOS Image Sensor with Pixel-Level Open-Loop Voltage Amplification," in ISSCC Dig. Tech. Papers, pp. 402-403, February 2011.

[Meynants 11] Meynants, G., Bogaerts, J., Wang, X., Vanhorebeel, G., "Backside Illuminated Global Shutter CMOS Image Sensors," Int. Image Sensor Workshop, Hokkaido, Japan, June. 2011.

[Mills 09] Mills, A. and Dudek, G., "Image stitching with dynamic elements," Image and Vision Computing, Volume 27, Issue 10, 2,2009.

[Nakamura 06] Nakamura, J., Image Sensors and Signal Processing for Digital Still Cameras. CRC Press, 2006.

[Nayar 06] Nayar. S., "Computational cameras: Redefining the image," Computer,vol. 39, no. 8, pp. 3038, 2006.

[Ohta 07] Ohta, J., Smart CMOS Image Sensors and Applications. CRC Press, 2007.

[Petschnigg 04] Petschnigg, G., Agrawala, M., Hoppe, H., Szeliski, R., Cohen, M.,

Toyama, K., "Digital photography with flash and no-flash image pairs," in ACM Trans. Graph. 23, 3, 2004, pp. 664672.

[Potmesil 83] Potmesil, M., Chakravarty, I., "Modeling motion blur in computer-generated images," Computer Graphics 17, 3 (July), 389-399, 1983.

[Raskar 04] Raskar, R., Tan, K., Feris, R., Yu, J., Turk, M., "Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging," in ACM Transactions on Graphics 23, 3 (Aug.), 679688, 2004.

[Raskar 06a] Raskar, R., "Computational Photography," Proc. Eurographics, 2006.

[Raskar 06b] Raskar, R., Agrawal, A., and Tumblin, J., " Coded exposure photography: Motion deblurring using fluttered shutter," ACM Transactions on Graphics 25, 3, 795804, 2006.

[Razavi 01] Razavi, B., Design of Analog CMOS Integrated Circuits. New York:McGraw-Hill, 2001.

[Reinhard 06] Reinhard, E., Ward, G., Pattanaik, S., Debevec, P., "High Dynamic Range Imaging: Acquisition, Display and Image-based Lighting," ISBN-10: 0-12-585263-0, Morgan Koffman Publishers, 2006.

[Stoppa 10] Stoppa, D., Massari, N., Pancheri, L., Malfatti, M., Perenzoni, M., and Gonzo, L., "An 80 60 range image sensor based on 10 m 50 MHz lock-in pixels in 0.18 m CMOS," in ISSCC Dig. Tech. Papers, pp. 406-407, February 2010.

[Sun 06] Sun, J., Li, Y., Kang, S. B., and Shum, H.-Y., "Flash matting." ACM Transactions on Graphics, 25(3):772778, 2006.

[Sun 07] Sun, J., Kang, S.-B. , Xu, Z., Tang, X., and Shum, H.-Y., "Flash cut: Foreground extraction with flash/no-flash image pairs.," in Proc. CVPR, 2007.

[Szeliski 10] Szeliski, R. Computer Vision: Algorithms and Applications. Springer, 2010.

[Takemura 03] Takemura, K., Oda, K., Nishimura, T. et al., "Challenge for Improving Image Quality of a Digital Still Camera," Proceedings of SPIE Vol.5017, pp. 385-392, Jan. 2003.

[Teranishi 84] Teranishi, N., Kohno, A., Ishihara, Y., Oda, E., and Arai, K., "An interline CCD image sensor with reduced image lag," IEEE Trans. Electron Devices, vol. ED-31, no. 12, pp. 1829-1833, Dec. 1984.

[Theuwissen 95] Theuwissen, A. J. P. Solid-state imaging with charge coupled devices. Kluwer Academic Publishers, Dordrecht; Boston, 1995.

[Theuwissen 06] Theuwissen, A. J. P., "The hole role in solid-state imagers," IEEE Trans. Electron Devices, vol. 53, no. 12, pp. 29722980, Dec. 2006.

[Toyama 11] Toyama, T., et al., "A 17.7Mpixel 120 fps CMOS image sensor with 34.8Gb/s readout," in ISSCC Dig. Tech. Papers, pp. 420-421, February 2011.

[Wan 11a] Wan, G., Li, X., Agranov, G., Levoy, M., Horowitz, M., "A Dual In-Pixel Memory CMOS Image Sensor for Computation Photography," Symp. VLSI Circuits Dig., June 2011.

[Wan 11b] Wan, G., Li, X., Agranov, G., Levoy, M., Horowitz, M., "CMOS Image Sensors with Multi-Bucket Pixels for Computational Photography," submitted to IEEE Journal of Solid-State Circuits, Sept. 2011.

[Wandell 02] Wandell, B. A., El Gamal, A., and Girod, B., "Common Principles of Image Acquisition Systems and Biological Vision", Proc. IEEE, vol. 90, 5-17, 2002.

[Wang 06] Wang, X., Rao, P. R., Mierop, A., Theuwisssen, A. J. P., "Random Telegraph Signal in CMOS Image Sensor Pixels," International Electron Device Meeting (IEDM06), pp. 1 4, Dec. 2006.

[Ward 03] Ward, G., "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," Journal of Graphics Tools, 8(2):1730, 2003.

[White 74] White, M. H., Lampe, D. R., Bllaha, F. C., and Mack, L. A., "Characterization of surface channel CCD image arrays at low light levels," IEEE J. Solid-State Circuits, SC-9(1), pp. 1-13, 1974.

[Xiao 01] Xiao, F., DiCarlo, J., Catrysse, P., and Wandell, B., "Image analysis using modulated light sources," in Proc. SPIE Electronic Imaging Conf., San Jose, CA, 2001, pp. 2230.

[Yamamoto 06] Yamamoto, K., Oya, Y., Kagawa, K., Nunoshita, M., Ohta, J., Watanabe, K., "A 128 128 Pixel Complementary Metal Oxide Semiconductor Image Sensor with an Improved Pixel Architecture for Detecting Modulated Light Signals," Opt. Rev. 13 (2006) 64.

[Yang 99] Yang, D. and El Gamal, A., "Comparative analysis of SNR for image

sensors with enhanced dynamic range," in Proc. SPIE, San Jose, CA, 1999, vol. 3649, pp. 197211, 1999.

[Yasutomi 10] Yasutomi, K., Itoh, S., Kawahito, S., "A 2.7e- temporal noise 99.7% shutter efficiency 92dB dynamic range CMOS image sensor with dual global shutter pixels," IEEE Int. Solid-State Circuits Conf.(ISSCC) 2010, Dig. Tech. Papers, pp398-399, San Francisco, Feb.2010.

[Yuan 07] Yuan, L., Sun, J., Quan, L., and Shum, H.-Y., "Blurred/noblurred image alignment using kernel sparseness prior," In IEEE International Conference on Computer Vision (ICCV), 2007.