

7.6 A 512×424 CMOS 3D Time-of-Flight Image Sensor with Multi-Frequency Photo-Demodulation up to 130MHz and 2GS/s ADC

Andrew Payne, Andy Daniel, Anik Mehta, Barry Thompson, Cyrus S. Bamji, Dane Snow, Hideaki Oshima, Larry Prather, Mike Fenton, Lou Kordus, Pat O'Connor, Rich McCauley, Sheetal Nayak, Sunil Acharya, Swati Mehta, Tamer Elkhatib, Thomas Meyer, Tod O'Dwyer, Travis Perry, Vei-Han Chan, Vincent Wong, Vishali Mogallapu, William Qian, Zhanping Xu

Microsoft, Mountain View, CA

Interest in 3D depth cameras has been piqued by the release of the Kinect motion sensor for the Xbox 360 gaming console [1,2,3]. This paper presents the pixel and 2GS/s signal paths in a state-of-the-art Time-of-Flight (ToF) sensor suitable for use in the latest Kinect sensor for Xbox One. ToF cameras determine the distance to objects by measuring the round trip travel time of an amplitude-modulated light from the source to the target and back to the camera at each pixel. ToF technology provides an accurate high pixel resolution, low motion blur, wide field of view (FoV), high dynamic range depth image as well as an ambient light invariant brightness image (active IR) that meets the highest quality requirements for 3D motion detection.

Depth and active IR images are produced by combining multiple images that are captured at different phase relationships of the clocks provided to the light source and pixel array. The captures are taken in rapid temporal succession to avoid motion blur. In addition, high differential dynamic range is necessary to simultaneously render high-reflectivity objects near the camera and low-reflectivity objects far from the camera. High dynamic range is realized by allowing each pixel to independently select the best shutter time (*multi-shutter*) and the best amplifier gain setting (*multi-gain*) at each capture.

Due to the multiple captures that need to be taken in rapid succession and the high dynamic range requirements, ADC conversion must be performed many times per capture and due to noise considerations cannot happen simultaneously with integration. Therefore a high-bandwidth 2GS/s 10b, column-parallel ADC is employed. Noise and mismatches are cancelled by using a completely differential design from pixel through ADC.

The ToF chip includes a 512×424 pixel array with 10 μ m pixel pitch fabricated in a standard TSMC 0.13 μ m CMOS LP 1P5M process. The 60% fill-factor (effective with μ Lens) pixel achieves a modulation contrast (MC) of 67% (measured at 50MHz) and a responsivity of 0.14A/W at 860nm. The chip can operate at high modulation frequencies of up to 130MHz to extract maximum depth quality while minimizing system light-source power. The schematic of the fully differential pixel design with a simplified detector plan is shown in Fig. 7.6.1. Capacitors C_{Int_A} and C_{Int_B} , are MIM caps and MSF_A & MSF_B are native source followers. Special care was taken in the pixel layout to maximize symmetry.

The pixel-timing diagram for an exemplary capture is also shown in Fig. 7.6.1. Correlated double sampling (CDS) is used to cancel the differential reset kT/C and fixed-pattern noises. During integration of a capture, ClkA & ClkB are modulated 180° out of phase for time t^{int} at the chosen modulation frequency and relative clock phase between the light source and pixel array. Exemplary integrated signals D_A & D_B are also shown in Fig. 7.6.1. At the end of integration, ClkA & ClkB are turned OFF and Read is turned ON to take an integration sample ($BitlineA^{int} - BitlineB^{int}$). If multi-shutter is used, one or more integrations with different exposure times (e.g., t^{int2}, t^{int3}) may follow before the pixel is Reset again for the next light phase/frequency. To further avoid pixel Common-Mode saturation in the presence of large amounts of ambient light, *Common-Mode-Reset* (CMR), which cancels common mode while preserving differential mode may be performed intermittently between integration cycles (as shown in Fig. 7.6.1) [4].

Figure 7.6.2 shows a cut view of the pixel detector when modulation clock ClkA (connected to poly-gates A shaped like fingers normal to the figure) is at ground and ClkB (connected to poly-gates B) is at a higher positive bias. Under these bias conditions, negative photo charges are collected under the gate oxide of the poly-gates. Upon collection under a poly gate, charges diffuse (in a direction normal to the plane of the figure) to a floating diffusion (FD) n+ collection node not in the plane of the figure. Potential barriers created by p+ doped areas between gates ensure that charges collected by one gate are never transferred to an adjacent gate even if it is at a higher potential.

Electric field lines from gates A (respectively B) shown in black (respectively white) define two distinct non-overlapping zones: a large ZoneA where the field lines terminate under gates A and a much smaller ZoneB for B. Notice that the electric field lines run tangent to the boundary between ZoneA and ZoneB (dotted line in Fig. 7.6.2). Photo charges generated in ZoneA are collected under gates A and similarly under gates B for ZoneB. Because ZoneA is much larger than ZoneB, most photo charges created by light arriving while ClkA is high are collected by gates A (and similarly for B). This assigns photo charges to either A or B depending on their arrival time with respect to ClkA and ClkB. The ratio (ZoneA-ZoneB)/(ZoneA+ZoneB) is approximately the pixel modulation contrast (MC) at low frequencies.

Since charges are never transferred between A and B, charge assignment to A or B occurs concurrently with charge collection under the gates and does not need an additional step of shifting charges between gates [3]. This makes our method (called: Quantum Efficiency Modulation) suitable for high modulation frequency. Ultimately, charges are collected at an FD node, but the timing is decoupled from charge assignment and can thus be performed *leisurely*.

The chip signal path starts with the 10 μ m-pitch differential column amplifiers shown in Fig. 7.6.3. First, the amplifier offset and pixel array column output voltages are sampled simultaneously onto 320fF input capacitors. Next, the input and feedback circuits are configured for high gain. Finally, the high gain output is compared with the ADC reference. If it would result in a saturated digital value at the output of the ADC, the amplifier is switched to a lower gain. This overflow event is reported to the system along with the ADC data.

The offset-cancellation switched-capacitor amplifier architecture allows programmable gain and provides better than 1% gain matching. It has a rail-to-rail input common-mode range and an output-referred offset that is nearly independent of gain. The amplifier also shifts the operating voltage from 3.3V to 1.5V to allow use of more efficient core transistors in the ADC.

The outputs from each group of 4 column amplifiers drive a 10b 8MS/s space-efficient 0.027mm² [5] successive-approximation ADC. A bank of 6 sampling capacitors capture the 4 column amplifier outputs on a round-robin basis. The complete 4:1 multiplexed converter is drawn on a 40 μ m pitch with an LSB capacitance of 4fF. 10b conversions with 8b ENOB complete every 10 clock cycles at 80MHz. A total of 256 ADCs on the chip produce over 2GS/s thus converting a full chip image capture in approximately 100 μ s.

The digitized output from the ADCs flows into the Shutter Engine, which choreographs CDS, multi-shutter, and multi-gain. The captured image is accumulated in an on-chip buffer and transferred via 8-lane 1Gb/s MIPI DPHY. The image capture rate is doubled by placing readout circuits at both the top and bottom of the pixel array.

Accuracy error of the system is generally better than 1% as presented in Fig. 7.6.4, which shows the mean measured distance vs. actual distance up to a range of 3.5m. This data is an average of 100 frames collected at 30fps where the reflectivity of the target was 10%. The standard deviation, also shown in Fig. 7.6.4, is 1cm at 3.5m with indoor fluorescent lighting and 1.5cm with 2.2 μ W/nm/cm². The high resolution, low motion blur, and wide FoV features of the sensor appear clearly in the captured depth and active IR images shown in Fig. 7.6.5. This ToF camera is capable of detecting up to 6 different persons in the FoV at a distance of about 3m. Figure 7.6.5 also shows a depth map generated at closer range that demonstrates the fine granularity of the depth data, which can detect not only individual fingers but wrinkles in clothing as well. A summary of important key metrics of this ToF sensor are listed in Fig. 7.6.6. Finally, a chip micrograph is presented in Fig. 7.6.7.

References:

- [1] C. Niclass, *et al.*, “A 0.18 μ m CMOS SoC for a 100m-Range 10fps 200×96-Pixel Time-of-Flight Depth Sensor”, *ISSCC Deg. Tech. Papers*, pp. 488-489, Feb. 2013.
- [2] L. Pancheri, *et al.*, “A QVGA-Range Image Sensor Based on Buried-Channel Demodulator Pixels in 0.18 μ m CMOS with Extended Dynamic Range”, *ISSCC Deg. Tech. Papers*, pp. 394-395, Feb. 2012.
- [3] W. Kim, *et al.*, “A 1.5MPixel RGBZ CMOS Image Sensor for Simultaneous Color and Range Image Capture,” *ISSCC Dig. Tech. Papers*, pp. 392-393, Feb. 2012.
- [4] C. Bamji, *et al.*, “Method and System to Differentially Enhance Sensor Dynamic Range” US. Patent 6,919,549 B2, July 2005.
- [5] Y. Suh, *et al.*, “A 10-bit 25-MS/s 1.25-mW Pipelined ADC With a Semidigital Gm-Based Amplifier,” *IEEE Trans Circuits and Systems II: Express Briefs*, Vol. 60, No. 3, pp. 142-146, Mar. 2013.

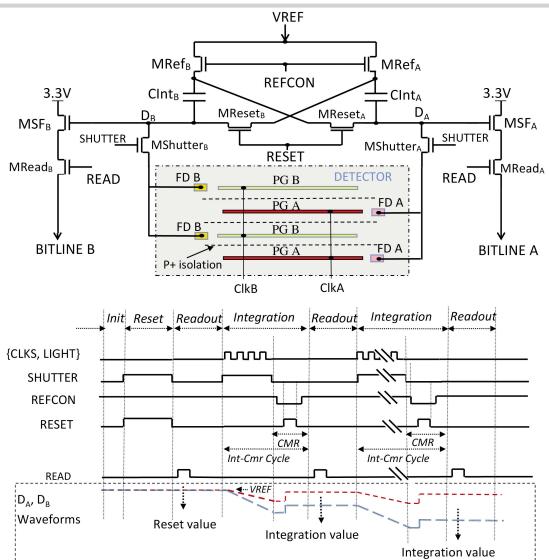


Figure 7.6.1: Pixel circuit and timing diagram.

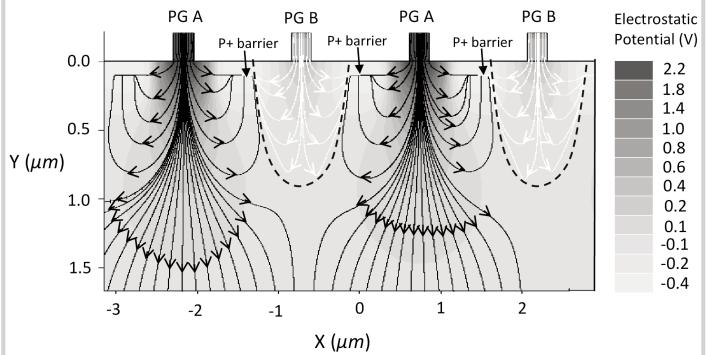


Figure 7.6.2: Detector device simulation.

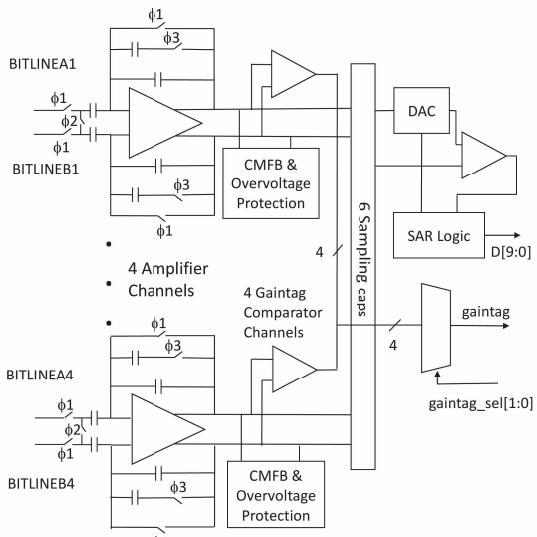


Figure 7.6.3: AMP and ADCs.

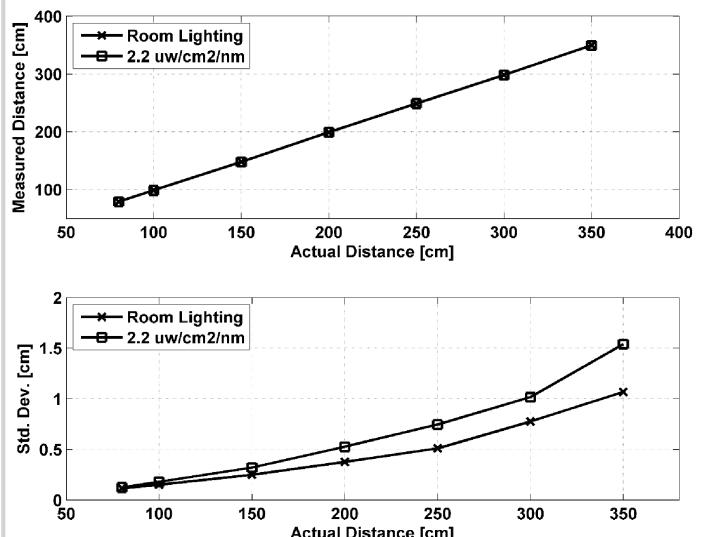


Figure 7.6.4: Accuracy and standard deviation.

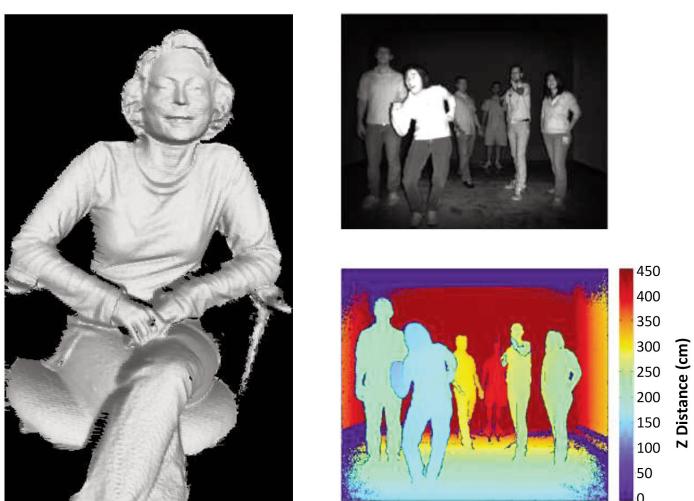


Figure 7.6.5: Chip data.

Process Technology	TSMC 0.13 1P5M
Pixel Pitch	10 μ m x 10 μ m
Pixel Array	512 x 424 Pixels
Chip size	8.2mm x 14.2mm
System Dynamic Range	> 2500 = 68db
Modulation Contrast	68% @ 860nm @ 50MHz
Modulation Frequency	10-130MHz
Average Modulation Frequency	80MHz
FOV	70 (H) x 60 (V) degrees
Depth Uncertainty	< 0.5% of range
Distance Range	0.8-4.2m
Operating Wavelength	860nm
Frame Rate	max 60fps (typical 30fps)
ADC	2GS/s
Effective Fill Factor	60%
Reflectivity	15%-95%
Chip Power	2.1W
Responsivity @ 860nm	0.144 A/W
Readout Noise	320 uV differential
F#	1.07
ADC Resolution	10

Figure 7.6.6: Performance parameters.

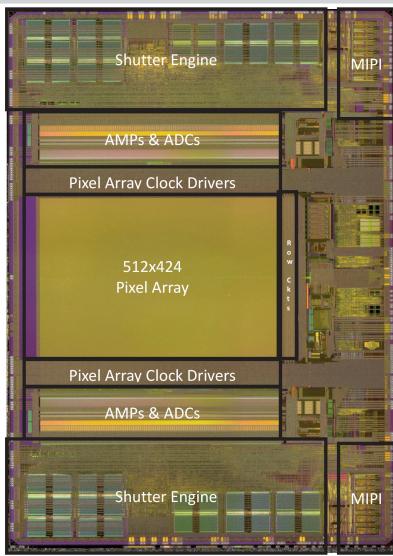


Figure 7.6.7: Chip micrograph.