

CEDAR: Computing-in-pixel Edge-aware Detection and Reconstruction Architecture for High-resolution 3D Imaging

Bu Chen^{1,2,*}, Zhangcheng Huang^{1,2,*†}, Qi Zheng¹, Weiyi Tang^{1,2}, Jingyi Wang^{1,2}, Hankun Lv^{1,2},

Chixiao Chen^{1,2}, Jianlu Wang^{1,2}, Qi Liu^{1,2,†}

¹State Key Lab of Integrated Chips and Systems, Fudan University, Shanghai, China

²Frontier Institute of Chip and System, Fudan University, Shanghai, China

ABSTRACT

Large-format single-photon avalanche diode (SPAD)-based direct time of flight (dToF) sensors are expected to be widely applied in future L5 full driving automation. However, the high-power in-pixel TDCs and the huge amount of data generated by multi-frame histogram sampling impose limitations on the pixel format of SPAD-based dToF sensors. To tackle this challenge, we proposed the Computing-in-pixel Edge-aware Detection and Reconstruction (CEDAR) architecture. In this architecture, edge pixels are recognized by charge-domain convolution (CDC) computing, and noise pixels are eliminated by in-memory denoising (IMD). Only few TDCs in these edge pixels are activated, resulting in significant power and data savings. Afterward, the full-format image is reconstructed by a U-Net using the obtained depth information from these edge pixels. For the first time, we proposed a high-resolution 512×512 SPAD-based dToF sensor with a low power of **83.3 mW**, a distance accuracy of **0.9 cm**, and a frame rate of **60 fps**. The high-resolution 3D image can be reconstructed by only 3.5% sparse edge pixels, achieving a PSNR of **35.2 dB**. The CEDAR architecture can achieve **16x** pixel format and image resolution improvement under the same constraint of power dissipation.

KEYWORDS

3D Imaging, Image Reconstruction, SPAD, Computing in Pixel

1 INTRODUCTION

Light detection and ranging (LiDAR) have been widely applied in autonomous driving, human face recognition, and biomedical imaging[1]. Direct time-of-flight (dToF) LiDAR receives pulsed lasers with high peak power and can detect targets in long distances, thus becoming a major solution in the field of autonomous driving. In the past, many LiDAR systems generated a 360-degree panoramic 3D image by rotating a linear-mode avalanche photodiode (LmAPD)

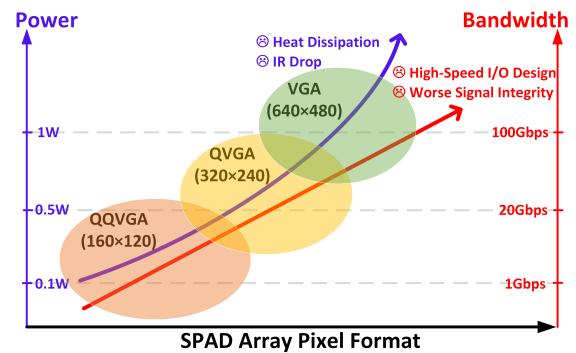


Figure 1: Power and bandwidth bottlenecks with SPAD array pixel format increasing.

array. Nowadays, more compact designs like prism-based, MEMS-based, and mirror-based structures are adopted in the LiDAR system, which reduces the volume, weight, and complexity[2].

Single-photon avalanche diode (SPAD) operating in Geiger mode, renowned for ultra-high gain and time resolution, is considered to be an alternative to LmAPDs. Flash dToF SPAD, which generates 3D imaging without any rotating structures, has attracted significant research interest and is considered to play an important role in the evolution of future L5 full driving automation. In the past two decades, CMOS-based SPAD technology has shown remarkable advancements in parameters such as photon detection efficiency (PDE), dark current rate (DCR), time resolution, and pixel density[3]. However, most SPADs are still in the pixel format of HQVGA, generating 3D images with low resolution. Some researchers operated the mega-pixel 2D-imaging SPAD in time-gating mode to generate high-resolution 3D image [4], but only at a short distance and ultra-slow frame rate. Currently, the imaging resolution of SPAD is much lower than that of CMOS image sensors (CIS), resulting in it not playing the expected role in autonomous driving and even having to be abandoned by some smart cars. The expansion of the SPAD pixel format faces significant challenges as follows:

Power Dissipation: The pixels of dToF SPAD record the time of flight of photons through a high-time-resolution, high-power time-to-digital converter (TDC). The power dissipation for an HQVGA-format SPAD detector can be as high as 2.5W[5]. The extremely high power consumption leads to a rapid increase in the temperature of the SPAD sensor and a deterioration in DCR, and introduces non-uniformity issues due to IR drop, even causing the SPAD sensor to fail. Some researchers have explored several solutions, such as reusing one TDC among multiple SPAD devices[6]. Nevertheless,

* Both authors contributed equally.

†Corresponding authors: huangzc, qi_liu@fudan.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DAC '24, June 23–27, 2024, San Francisco, CA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0601-1/24/06...\$15.00

<https://doi.org/10.1145/3649329.3657313>

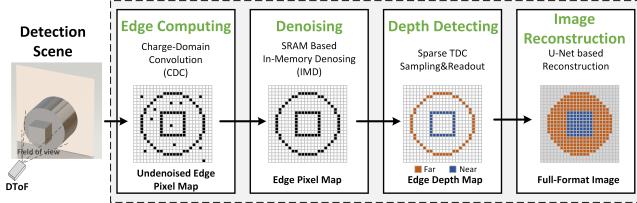


Figure 2: Concept of the CEDAR architecture

this approach results in decreased array resolution and deteriorates signal-to-noise ratio (SNR) due to photon loss.

Data Storage: The data generated by SPAD originate not only from the reflected laser pulses but also from background light and DCR. Multiple frames, for example, 100 to 10000, are needed for histogram statistics to extract the flight time of laser pulses from the peak position. For large-format SPAD arrays, this results in tremendous storage overhead. Approaches like sliding[7] and zooming[8] have been employed to compress histograms and reduce storage, but these methods sacrifice the detection frame rate.

IO Bandwidth: With the expansion of SPAD pixel format, the volume of output data experiences a substantial increase. As shown in Fig.1, for a VGA-format SPAD sensor operating at a 30fps frame rate, a high IO bandwidth of 120 Gbps is needed to transmit the full array of data, putting huge pressure on the number of IOs and power consumption. Some researchers utilized more efficient asynchronous readout[9] instead of synchronous readout, but only for sparse dataflow scenarios. On-chip histograms can drastically reduce the amount of data that is transmitted out of the sensor chip, but they require a large amount of storage space and do not reduce the amount of data at the front end of the SPAD device.

To address the aforementioned challenges, we first propose a computing-in-pixel edge-aware detection and reconstruction (**CEDAR**) architecture for high-resolution 3D imaging. Moreover, we designed a 512×512 dToF SPAD detector utilizing the CEDAR architecture. Our primary contributions are as follows:

(1) Our proposed CEDAR is a reconstructive 3D imaging architecture based on sparse features, where only the depth edges in the scene are detected by the SPAD sensor, thus the architecture greatly reduces the power and data bandwidth of the SPAD sensor.

(2) We designed a 512×512 SPAD sensor based on CEDAR architecture. The in-pixel charge-domain calculation and denoising are able to distinguish edge pixels in the array. This allows the chip to achieve 3D imaging at a high frame rate and low power dissipation.

(3) We used a U-Net to realize the high-resolution 3D image. Based on the sparse depth information of edge pixels, full-format 3D depth images with high PSNR on average can be reconstructed.

2 BACKGROUND AND RELATED WORK

2.1 Elements of the SPAD-based dToF Sensors

- **Pulsed Laser:** The laser emits laser pulses with a narrow width at a fixed interval to illuminate the target.
- **Quenching Circuit:** The quenching circuit promptly releases the anode charge to reset the SPAD after an avalanche triggered by a photon, effectively reducing the dead time of the SPAD.

- **Time-to-Digital Converter:** The TDC records the time interval between the laser emission and the detection of a photon. Typically, a TDC consists of a high-precision ring oscillator (RO) and a ripple counter[10].

- **Histogram:** A histogram is employed to record the distribution of TDC code values across multiple frames, mitigating interference from non-ideal factors like ambient light and clock jitter. The peak position in the histogram typically corresponds to the actual depth position of the target.

2.2 Processing Algorithms for 3D Image Sensors

- **Edge-Detection:** Edge detection aims to identify contours in the captured image where pixel depth exhibits significant variation when compared with neighboring pixels. The recognition of these pixels is computed by edge-detection operators including Sobel, Robert, Laplacian, and Canny operators, etc.
- **Image Reconstruction:** The algorithm learns the relationship between partial pixels and full pixel array, then converts the sparse depth data of partial pixels into full array images. Tasks associated with image reconstruction encompass image super-resolution[11], high dynamic range (HDR) reconstruction[12], frame interpolation[13], etc. Neural network models are frequently employed for image reconstruction owing to their powerful feature extraction capabilities. U-Net excels in image reconstruction with its encoder-decoder architecture and skip connections, effectively preserving both local and global features, which is particularly well-suited for tasks related to image reconstruction[14].

3 MOTIVATION

Currently, the image resolution of SPADs cannot meet the requirements for fully autonomous driving, prompting the exploration of new methods for developing large-format SPADs. If only few pixels in the edge of the depth image are activated and generate the depth information, the power consumption of the pixel array and data to be transferred will be significantly reduced. Therefore, the pixel format of SPAD can greatly scale up because the power consumption and bandwidth constraints are broken through. Full-format 3D images can be reconstructed by neural network-based algorithms based on the few depth information in these edge pixels.

Based on this concept, we propose the CEDAR architecture, as shown in Fig.2. Firstly, the SPAD sensor operates in edge computing. The original time of flight captured by SPADs, along with the in-pixel weights, are computed using energy-efficient analog-domain convolution. Pixels on the depth edge can be recognized by the convolution results. Next, the SPAD sensor carries out the denoising. Scattered noise pixels are eliminated by average filtering and time-domain noise caused by background light is reduced by repeating the above edge computing multiple times. Then, the SPAD sensor operates in depth-detecting, activating only the previously identified edge contour pixels to detect their depth information. Finally, an off-chip U-Net network is utilized to restore the entire 3D depth image from the obtained sparse edge depth.

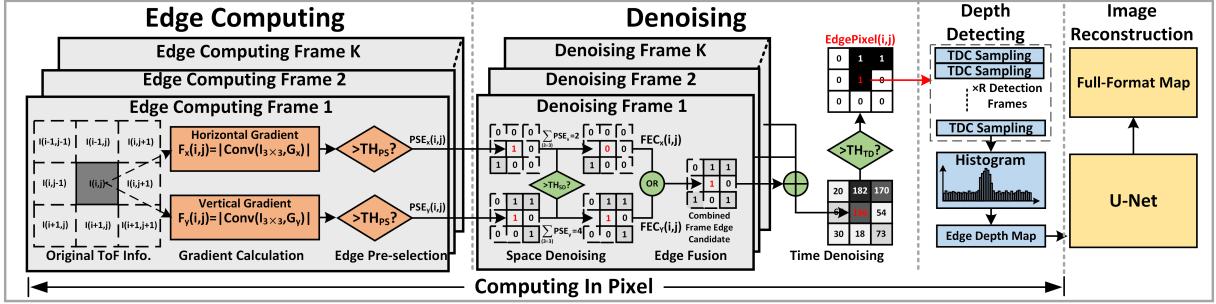


Figure 3: Block diagram of the CEDAR.

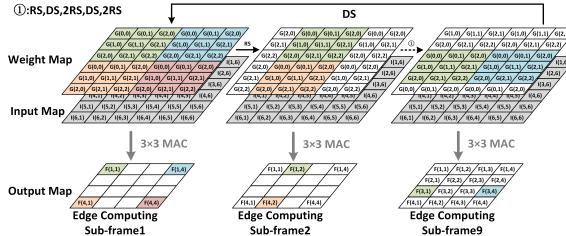


Figure 4: Weight-shifting mode convolution.

4 DESIGN OF CEDAR

Fig.3 illustrates the overview of the CEDAR architecture, which includes blocks of edge computing, denoising, depth detecting and image reconstruction.

4.1 Edge Computing

A pixel can be recognized to be on the depth edge in the 3D scene if its time of flight is significantly different from its neighbors. We

$$\text{use the Sobel operators } (G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix})$$

to extract the edge pixels owing to its excellent noise-smoothing capability. The weight-shifting mode convolution is shown as Fig.4. By taking the time of flight as the input map and the Sobel operators as the weight map, we carry out convolution computing and obtain the gradient magnitudes in both vertical and horizontal directions.

In the next sub-frame of edge computing, the entire weight map and kernel window execute right-shifting (RS) and down-shifting (DS) for the neighbor pixel in horizontal (x) and vertical (y) directions, respectively. For a 3×3 kernel window, the entire output map F_x and F_y can be obtained after completing only 9 times of shifting, as follows:

$$F_{x,y}(i,j) = \left| \sum_{m,n \in [-1,1]} I(i+m, j+n) \cdot G_{x,y}(m+1, n+1) \right| \quad (1)$$

The output map of the gradient convolution represents the difference in depth between the pixel and the surrounding pixels. TH_{PS} is used to generate pre-selected edge (PSE) pixels as shown in Eq.2.

$$PSE_{x,y}(i,j) = \begin{cases} 1 & F_{x,y}(i,j) > TH_{PS} \\ 0 & \text{else} \end{cases} \quad (2)$$

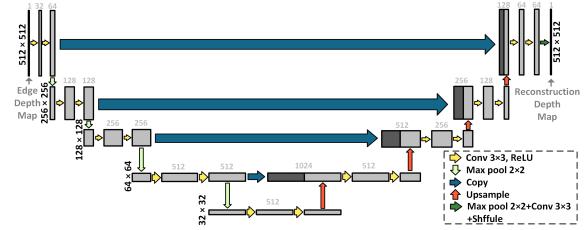


Figure 5: U-Net structure for CEDAR image reconstruction.

4.2 Denoising

The pre-selected edge pixels contain several noise pixels due to ambient light and DCR. In denoising computation, there are three operations including spatial denoising, edge fusion, and temporal denoising. As shown in Fig.3 and Eq.3, if the number of PSE pixels in the 3×3 region around $PSE_{x,y}(i,j)$ is greater than TH_{SD} , the pixel, (i,j) , is considered to be the frame edge candidate (FEC) pixel and is set to '1'. Edge fusion combines the frame edge candidate pixels in the x and y directions. Only spatial denoising is not enough to reduce noise from ambient light and DCR. Therefore, we count the number of combined FEC in K frames for each pixel. Noise pixels can be eliminated when compared with the threshold, TH_{TD} , as shown in Eq.4.

$$FEC_{x,y}(i,j) = \begin{cases} 1 & \sum_{m,n \in [-1,1]} PSE_{x,y}(i+m, j+n) > TH_{SD} \\ 0 & \text{else} \end{cases} \quad (3)$$

$$EdgePixel(i,j) = \begin{cases} 1 & \sum_K FEC_x(i,j) \cup FEC_y(i,j) > TH_{TD} \\ 0 & \text{else} \end{cases} \quad (4)$$

4.3 Depth Detecting

Edge pixels are recognized and labeled in the SPAD array according to the above operations. Only edge pixels are activated and other pixels are ceased. To obtain the accurate depth of these edge pixels, depth detecting is carried out in R times to generate a histogram. The true depth is abstracted from the peak position in the histogram.

4.4 Image Reconstruction

Sparse depth data of edge pixels are sent to a U-Net and generates a 512×512 input map by filling zero in non-edge pixels. The U-Net has down-sampling, up-sampling, 3×3 convolution, 2×2 max

pooling, and other operations. A full 512×512 depth image is produced by this U-Net as shown in Fig.5.

5 THE SPAD-BASED DTOF SENSOR WITH CEDAR ARCHITECTURE

5.1 Chip Overview

The overall structure of the SPAD sensor with the function of computing-in-pixel edge-aware detection is shown in Fig.6. The pixel consists of a charge-domain convolution (CDC) circuit, an in-memory denoising (IMD) circuit, a TDC, and others. 512 column-level sparse data buffers are used to compress data from the pixel array. A PLL, a clock tree, an MCU, two weight buffers, and a readout circuit are included in the SPAD sensor.

5.2 CDC Circuit

The CDC circuit is used to find out the edge pixel preliminarily.

(1) **SPAD Front-End:** Photon is detected by the SPAD device and a pulse with a width equal to the time of flight is generated. This wide pulse corresponds to the input in the convolution computation.

(2) **Weight Shifting and Gating Circuit:** When in weight-shifting mode, the weights are stored as 4-bit signed data in the D flip-flops. At the end of one computation cycle, the weights are shifted in the horizontal or vertical direction which is controlled by Weight_Sel. When in gating mode, the weight value W[2:0] and Original_ToF generate the current source gating signal GT[2:0], and the MSB and MSB_N act as selection signals of the current source set.

(3) **Gradient Calculation Circuit:** It includes two sets of current sources, a set of differential capacitors, C_P and C_N , and several switches. At the beginning, the COMP switch is disconnected and both capacitors are pre-charged to VDD. Whether the capacitor is discharged is determined by the MSB. GT[2:0] generates the weighted current with a ratio of 4:2:1 and controls the integration time. The integration voltage on the capacitance is proportional to the multiplication of the inputs and weights. To sum the multiplication result, nodes ② and ③, in the same kernel window, will be connected by the switches. The voltage of capacitors is shown in Eq.5.

$$V_{C_{P(N)}} = VDD - \frac{I_0}{9C_{P(N)}} \sum_{3 \times 3} MSB(-N) \cdot (4GT[2] + 2GT[1] + GT[0]) \quad (5)$$

Finally, we turn on the COMP switch, and according to charge conservation, we can get a voltage, V_G , that is linearly related to the convolution computing as shown in Eq.6.

$$V_G = \frac{1}{2}VDD + \frac{1}{2}(V_{C_P} - V_{C_N}) \quad (6)$$

The simulation results of the gradient computation circuit are shown in Fig. 7. The linearity of CDC output reaches 99.8% in 1000 times of 3×3 random convolution experiments.

(4) **Edge Pre-Selection Circuits:** The gradient calculation result, V_G , is compared with two comparators. If V_G is located outside the window of two thresholds, this pixel will be regarded as the pre-selected edge (PSE) pixel and is labeled to be '1' in the PSE latch.

5.3 IMD Circuit

The IMD circuit includes two spatial denoising and a temporal denoising circuit which aims to reduce the noise in pre-selected pixels.

(1) **Spatial Denoising Circuit:** After completing the edge pre-selection calculation for all pixels through 9 edge computing frames, the sampling switch DE_S and the computation switch DE_C are turned on and the values in the PSE latch are sampled into a capacitor, C_{PSE} . After turning off DE_S, the switch DE_S_N will connect the C_{PSE} of the current denoising pixel and its surrounding 8-pixel ④. After charge redistribution, the voltage of the C_{PSE} can be obtained as:

$$V_{C_{PSE}} = \frac{VDD}{9} \sum_{m,n \in [-1,1]} PSE(i-m, j-n) \quad (7)$$

After turning on the feedback switch DE_C_N, the positive feedback in the 6T-SRAM restores the output voltage to '0' or '1' and rewrites the latches in the selected PSE pixels. V_{SDTH} adjusts the denoising threshold in the range from 2 to 6.

(2) **Temporal Denoising Circuit:** The spatial denoising results in both vertical and horizontal directions, DE_RESULT_V(H), are combined by the OR operator. The capacitor, C_{TD} , counts multiple spatial denoising results as an analog counter. Based on the threshold, V_{TDTH} , the comparator outputs the final calculation result EDGE_FLAG.

5.4 In-Pixel TDC and Peripheral Circuit

Each pixel contains a 14-bit TDC with a time resolution of 60 ps, corresponding to a distance accuracy of 0.9 cm. EDGE_FLAG is used to activate TDCs in edge pixels in detecting frames. The data in activated pixels are sent to a sparse data buffer (SDB) in each column. The buffer contains a FIFO with a depth of 5 and encodes the pixel information into 24b. The readout circuit of the chip includes data multiplexers, parallel-in-serial-out (PISO), and 6 LVDS IOs at a rate of 400Mbps. In addition, the chip also includes a row and column weight buffer for writing weight into SPAD arrays, respectively. PLL and clock tree are adopted to generate clocks for detection, and MCU is used for global control, etc.

5.5 Working Flow of the SPAD Sensor

The working flow of the SPAD sensor is shown in Fig.8, demonstrating the ability of high frame rate detection. The repetition frequency of the laser is 1MHz. In each imaging frame, the time of edge map computing is 2ms and 100 times of edge map detecting and data readouts occupies 12.9ms. The total imaging frame time is 16.6ms, achieving the 3D imaging at 60fps.

6 EXPERIMENTS

6.1 Experiments Setup

(1) **Dataset Augmentation:** We established car models from ModelNet40[15] along with data augmentation. The car underwent horizontal movement within a depth range of 5-15m across a 30m scene. The depth maps for the ideal scene were generated with a Field of View (FOV) of 45°. Additionally, we utilized the depth maps from the SUN RGB-D[16] dataset captured by the Kinect2 sensor,

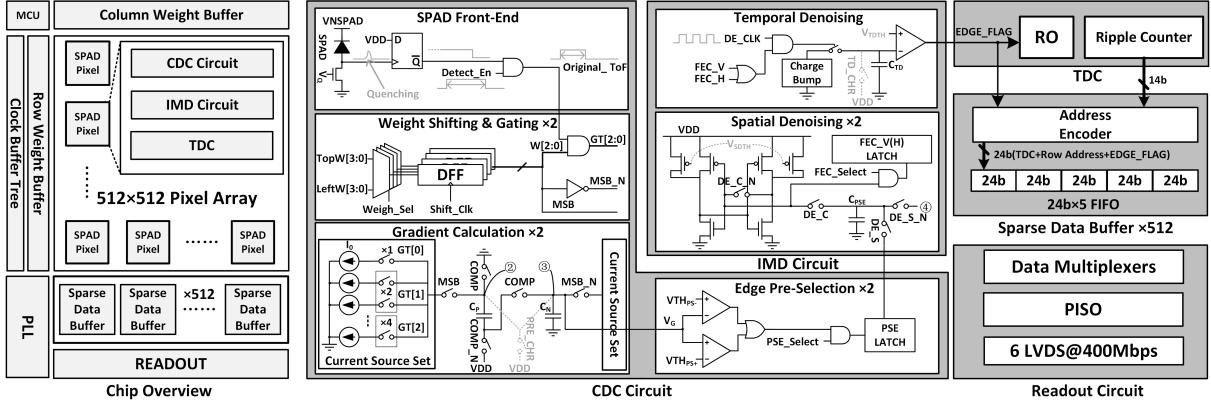


Figure 6: The block diagram and schematic of the SPAD sensor.

cropped to a resolution of 512×512 , to serve as the ground truth maps.

(2) **U-Net Training:** The training dataset and test dataset of ModelNet40 include 1379 images and 700 images, and SUN RGB-D includes 2700 images and 786 images, respectively. We trained the U-Net using one NVIDIA TITAN RTX GPU, setting the batch size to 18 and the training epoch to 100. We used L2 loss and Adam optimizer to optimize the network. The initial learning rate is $4e-4$, reduced by half every 20 epochs.

(3) **Behavioral Level Platform:** We established a Python-based behavioral simulation model for the sensor, comprising two components: the SPAD sensor and the U-Net. The SPAD sensor based on CEDAR samples the augmented dataset and produces edge depth maps. Subsequently, the U-Net conducts inference on these edge depth maps, completing the image reconstruction process.

(4) **Hardware Simulation:** We have designed the SPAD sensor using a 65nm CMOS process and simulated it to evaluate parameters such as area, power dissipation, and bandwidth.

6.2 Image Quality Analysis of CEDAR

Fig.9 demonstrates the edge pixel extracted by the CEDAR architecture on the Modelnet dataset. The percentage of the edge pixels in pixel arrays is only 0.56%. It means that CEDAR can greatly reduce power and bandwidth in simple scenarios.

Fig.10 shows that CEDAR also has excellent performance in complex scenarios. In the SUN RGB-D, the percentage of the edge pixels extracted by CEDAR is about 3.51%. A high-quality 3D image is reconstructed with an average PSNR of 35.1 dB.

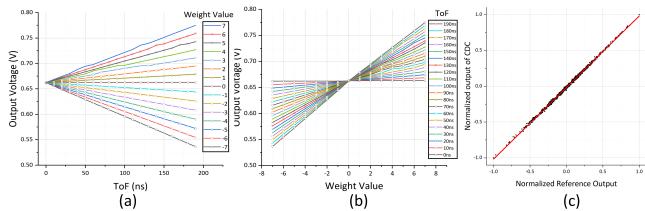


Figure 7: Simulation results of CDC: (a) Output voltage v.s. time of flight. (b) Output voltage v.s. weight. (c) Output linearity with 1000 random convolution tests.

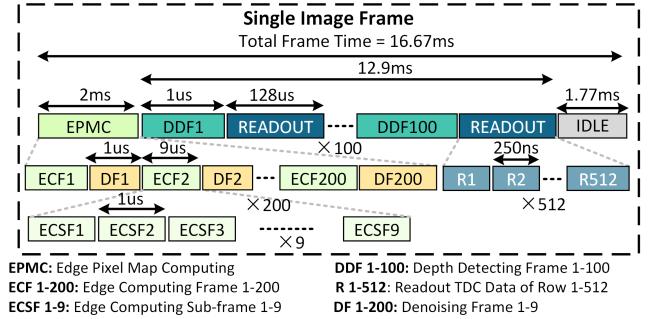


Figure 8: Workflow of the proposed chip for 3D imaging.

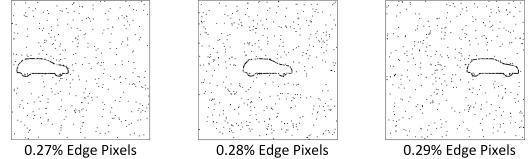


Figure 9: Edge pixels extraction in Modelnet40 by CEDAR.

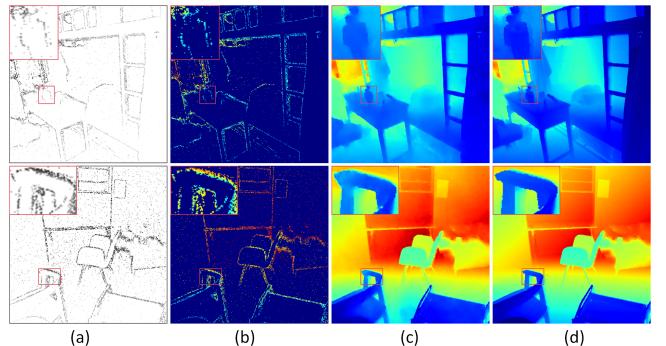


Figure 10: Visual results of (a) Edge pixel map, (b) Edge depth map, (c) Reconstruction map, and (d) Ground truth map for SUN RGB-D by CEDAR.

6.3 Power and Bandwidth Analysis of CEDAR

We compare the CEDAR architecture with traditional full-format detection at various pixel formats as shown in Fig.11. The power consumption of a 512×512 dToF SPAD sensor using the CEDAR is

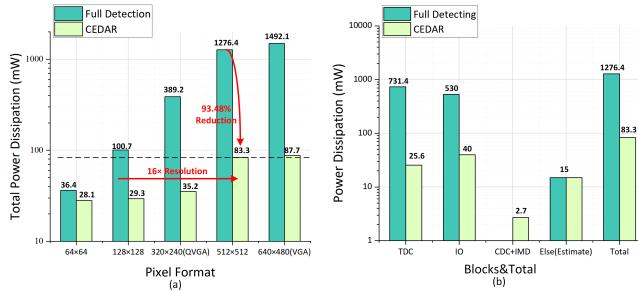


Figure 11: Comparison between CEDAR and traditional detection: (a) total power dissipation at different pixel format, (b) power dissipation of blocks and whole chip.

83.3mW, which is only 6.52% of the full-format detection. Therefore, by replacing full-format detection with the CEDAR, the pixel format can be augmented by more than 16 times under the same constraint of power dissipation. Fig.12 compares the power consumption of different blocks between CEDAR and traditional detection. At the same resolution, CEDAR lowers the TDC power consumption from 731 mW to 25.6 mW, and reduces the IO power from 530 mW to 40 mW. For CDC and IMD, power consumption is as low as 2.7mW which comes from capacitor charging. Table 1 shows the performance comparison of this work with other state-of-the-art (SOTA) dToFs. The dToF sensor using CEDAR firstly achieves high resolution and high frame rate 3D imaging within 100mW.

Table 1: Comparison of Performance with SOTA DToF SPAD Sensors

	[5]	[17]	[4]	[18]	This Work
Process	180nm	40nm/90nm	180nm	90nm/40nm	65nm
Detection Mode	Flash	Flash	Flash	Scanning	CEDAR
SPAD Pixel Format	252 × 144	256 × 256	1024 × 1000	189 × 600	512 × 512
3D Image Resolution	252 × 144	64 × 64	1024 × 1000	168 × 63	512 × 512
3D Image Frame Rate(fps)	30	30	<1	20	60
Power(mW)	2540	77.6	284/535	1192	83.3
Depth Time Resolution(ps)	48.8	35/560	36	1000	60
Laser Repetition Rate(MHz)	40	1.9	40	-	1
Pixel Pitch(μm)	28.5	9.2/38.4	9.4	10	30
Fill Factor(%)	28	51	7.0/13.4	-	40

7 CONCLUSION

We propose a new CEDAR architecture by computing-in-pixel edge-aware detection. In this architecture, only few edge pixels are activated, and sparse depth data are transferred, resulting in a significant reduction of power dissipation, storage, and bandwidth. The full-format 3D image can be reconstructed by U-Net using only 3.5% sparse depth information and shows a high quality with a PSNR of 35.2 dB. By the CEDAR, the 512×512 SPAD-based dToF sensor achieves a high frame rate of 60 fps by only 83.3 mW low power. The CEDAR exhibits excellent ability in 3D depth imaging and paves the way for future full driving automation.

ACKNOWLEDGEMENT

This work is supported in part by the National Natural Science Foundation of China under Grant No.62374039 and No.62235009, in part by the Shanghai Committee of Science and Technology under Grant No.21TS1401400.

REFERENCES

- [1] Li et al. Lida for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine*, 37(4):50–61, 2020.
- [2] Campbell et al. Recent advances in avalanche photodiodes. *Journal of Lightwave Technology*, 34(2):278–285, 2015.
- [3] Morimoto et al. 3.2 megapixel 3d-stacked charge focusing spad for low-light imaging and depth sensing. In *2021 IEEE International Electron Devices Meeting (IEDM)*, pages 20.2.1–20.2.4, 2021.
- [4] Kazuhiro Morimoto et al. Megapixel time-gated spad image sensor for 2d and 3d imaging applications. *Optica*, 7(4):346–354, Apr 2020.
- [5] Zhang et al. A 30-frames/s, 252 × 144 spad flash lidar with 1728 dual-clock 48.8-ps tdc, and pixel-wise integrated histogramming. *IEEE Journal of Solid-State Circuits*, 54(4):1137–1151, 2019.
- [6] Padmanabhan et al. 7.4 a 256×128 3d-stacked (45nm) spad flash lidar with 7-level coincidence detection and progressive gating for 100m range and 10klux background light. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, volume 64, pages 111–113, 2021.
- [7] Stoppa et al. A reconfigurable qvga/q3vga direct time-of-flight 3d imaging system with on-chip depth-map computation in 45/40 nm 3d-stacked bsi spad cmos. In *Proc. Int. Image Sensor Workshop*, pages 53–56, 2021.
- [8] Erdogan et al. A 16.5 giga events/s 1024 × 8 spad line sensor with per-pixel zoomable 50ps-6.4 ns/bin histogramming tdc. In *2017 Symposium on VLSI Circuits*, pages C292–C293. IEEE, 2017.
- [9] Mauro Buttafava et al. Spad-based asynchronous-readout array detectors for image-scanning microscopy. *Optica*, 7(7):755–765, Jul 2020.
- [10] Cui et al. Toward implementing multichannels, ring-oscillator-based, vernier time-to-digital converter in fpgas: Key design points and construction method. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 1(5):391–399, 2017.
- [11] M. et al. Elad: Super-resolution reconstruction of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):817–834, 1999.
- [12] Elertsen et al. Hdr image reconstruction from a single exposure using deep cnns. *ACM Trans. Graph.*, 36(6), nov 2017.
- [13] Zitnick et al. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, 23(3):600–608, aug 2004.
- [14] Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, pages 234–241. Springer, 2015.
- [15] Zhirong et al. Wu. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [16] Song et al. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [17] Henderson et al. 5.7 a 256×256 40nm/90nm cmos 3d-stacked 120db dynamic-range reconfigurable time-resolved spad imager. In *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, pages 106–108, 2019.
- [18] Kumagai et al. 7.3 a 189×600 back-illuminated stacked spad direct time-of-flight depth sensor for automotive lidar systems. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, volume 64, pages 110–112, 2021.