A Review of CMOS Image Sensor Pixel Structure Research

Huaxuan Yao
Shanghai Maritime University, Shanghai, China
SpeysideYHX@outlook.com

ABSTRACT

A CMOS image sensor (CIS) comprises millions of independent micro-units—known as pixel units—that convert light signals into electrical signals. Whilst its structure evolves rapidly, its adaptability remains relatively limited when faced with certain extreme conditions. Today, the primary question is what approach should be taken in future research to enable pixel structure to deliver better performance. This paper summarized current technologies by examining cutting-edge pixel structure designs and their technical evolution, whilst also forecasting future technological development trends. Through research methods such as product and technology comparisons, analysis of the drivers of technological evolution, and exploration of the gap between technical limits and physical limits, the paper identifies the technical bottlenecks and applicable scenarios for different structures, and proposes future research directions based on AI-driven collaborative improvement and quantum dot integration.

Keywords: Pixel structure; CMOS image sensor; High dynamic range; Pixel units; Three dimensional integrative circuits; Noise reduction.

1. INTRODUCTION

Nowadays, with the shift in research focus from ‘pixel minimization’ to ‘in-pixel intelligence’, we are witnessing the emergence of a host of advanced pixel structures. In addition to the 5T and 6T structures made by adding transistors to APS 3T and 4T, many companies and university teams have also proposed very new structures. For example, the Charge-Splitting High Dynamic Range 4-Tap Single-Exposure Sensor [1] has reached the dynamic range (DR) of 110dB, and the 60 dB ST BrightSense series of sensors [2] which greatly exceeds the traditional limit contains two modes of rolling shutter and global shutter in one sensor, and can switch modes one by one. Sony introduced the sensor of LYTIA 901[3], which used the AI-learning array reorganization conforming to QQBC arrangement and put it into the sensor. This AI-learning array reorganization can shoot small parts such as complex patterns and characters very beautifully. In the field of household appliances and industrial inspection, the development and innovation of pixel structure is valuable, so this paper divides this topic into four parts for research: 1. Tracing back the historical evolution of the pixel structure, analyzing the shortcomings of the old pixel structure and the new technological progress; 2. Analyzing the research progress and new technological measures of the pixel structure with good performance; 3. The latest attempt and use of the new special pixel structure are simple and easy. paper summarizes the current research on the pixel structure by comparing products and technologies, analyzing the reasons for technological progress, finding the gap between technical limits and physical limits, and simply points out the future research direction.

2. THE THEORETICAL BASIS AND EARLY DEVELOPMENT OF CMOS PIXEL STRUCTURE

This section analyzes traditional CMOS pixel structures based on their historical evolution and theoretical foundations.

2.1. The theoretical origins of traditional pixel structure

The theoretical basis of CMOS pixel structure can be traced back to the 1960s. In 1963, Morrison first proposed the concept of a semiconductor-based computable sensor structure. In 1967-1968, Fairchild Semiconductor’s Weckler system demonstrated the feasibility of reverse-biased photodiodes operating in integration mode, establishing the physical basis for photogenerated charge storage and readout, which became the foundation of all CMOS pixels [4]. Based on this, passive pixels (PPS) were developed, consisting only of photodiodes and MOS switches. The structure was extremely simple, but the signal attenuation was severe and the noise was high, proving the possibility of on-chip integration but making it difficult to put into practical use.

In the early 1990s, in order to overcome the signal-to-noise ratio bottleneck of PPS, the theory of CMOS active pixel sensors (APS) emerged. In 1992-1993, the Fossum team at NASA's Jet Propulsion Laboratory proposed integrating source followers within the pixel for signal amplification, forming a 3T-APS structure consisting of three transistors for reset, amplification, and row selection, thereby achieving low noise, low power consumption, and on-chip random access [5]. In April 1993, JPL successfully developed the first APS verification chip with 28×28 pixels and a center pitch of 40 μm, marking the beginning of the active pixel era for CMOS image sensors.

2.2. Literature review on classical structure

Weckler's paper on integral models published at the 1967 ISSCC, along with Reticon's linear array products, are the most frequently cited foundational documents in the PPS phase. The key documents that truly initiated the integration of APS academia and industry were concentrated in the mid-1990s. In 1994, the paper on CMOS Active Pixel Image Sensor of Mendis, Kemeny and Fossum explained the structure, working principle and noise model of 3T-APS in detail, which became a representative document in this field [6]. In 1995, Fossum further put forward the idea of "Camera-on-a-Chip". He said that all camera functions can be put in one chip, which promoted the changes of the whole industry [7].

In the industrial world, Fossum's team made Photobit in 1995. They launched 3T-APS, which confirmed that ordinary people can also use it. In 1997, Toshiba showed the 0.6 μm CMOS APS with 330,000 pixels on ISSCC, and the mechanism began to be produced on ordinary production lines. At the same time, researchers noticed that 3T pixels had a large fixed-mode noise. This is because of the deviation of kT/C noise and source follower of reset transistor. After that, a 4T-APS pixel with transfer gate and floating diffusion node appeared, and its noise was reduced by correlated double sampling, and the dark current and sensitivity were also improved [8]. Fossum team conducted a comparative experiment between 3T and 4T pixels, unified the fill factor, quantum efficiency, conversion gain and other indicators into an evaluation method, and developed the mainstream of high-performance pixels into 4 t-APS [8].

From Weckler's integration mode to the development of 3T/4T active pixels, CMOS pixel structures evolved from theoretical concepts into mature imaging systems. This evolution laid the theoretical and technical foundations for subsequent architectures, such as back-side illuminated and stacked systems. The major milestones in this development are summarized in Figure 1.

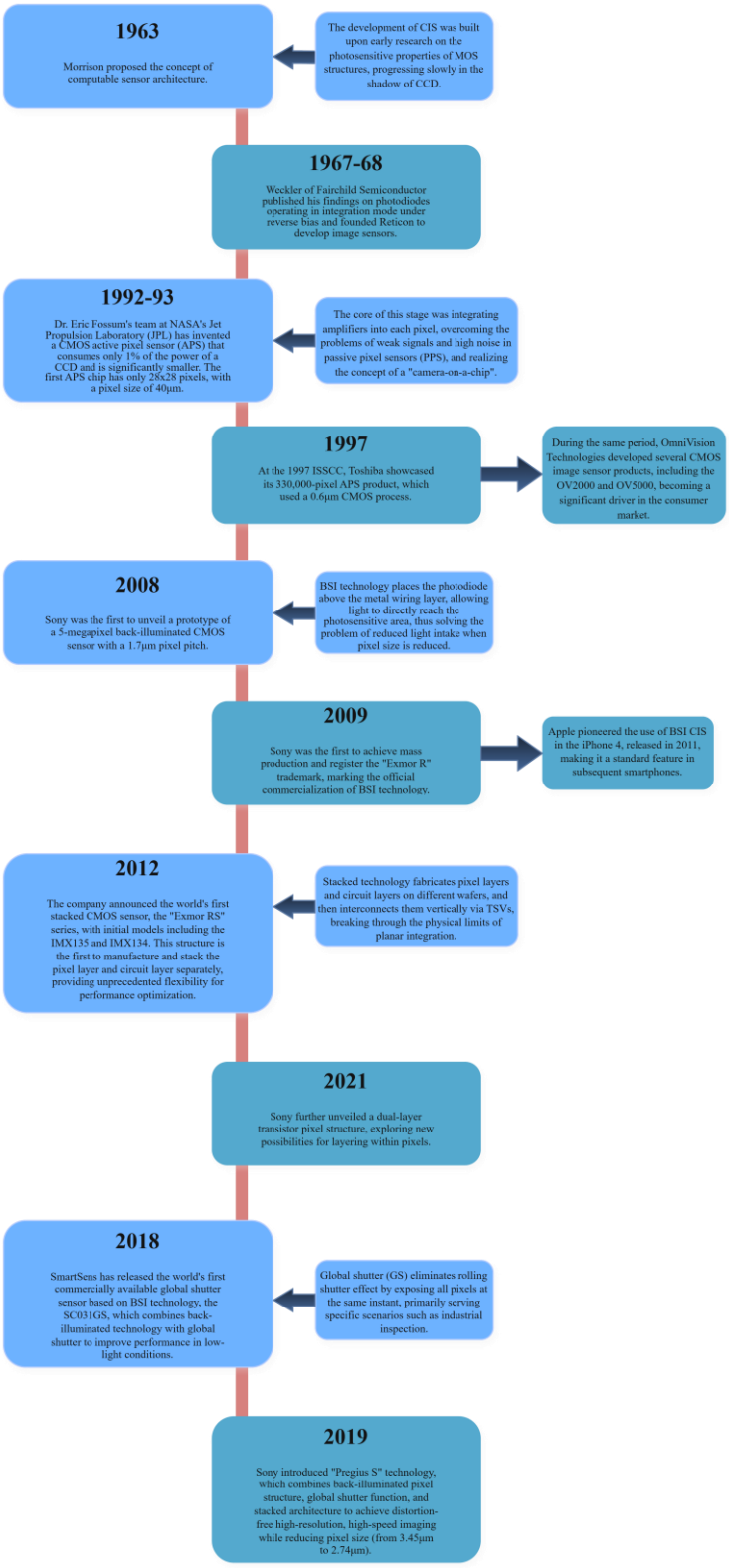


Figure 1. The historical evolution of CMOS image sensor pixel structure.

3. INNOVATIVE RESEARCH PROGRESS ON HIGH-PERFORMANCE PIXEL STRUCTURE

3T and 4T-APS create the basic working mechanism of CMOS image sensors. However, in the adjustment of sensitivity and noise, it has gradually reached the physical limit. Since the beginning of the 21st century, academia and industry have made many innovations in pixel structure. This chapter systematically organizes and explains the important research progress from two directions: high sensitivity pixel structure and low noise pixel structure.

3.1. Research on high sensitivity pixel structure

Traditional front-side illumination (FSI) is the first pixel structure used in CMOS image sensors. In the FSI structure, the photodiode is located on the front side of the silicon substrate and covered by the back-end fabrication structure composed of metal and dielectric layers. Incident light must pass through these opaque wiring layers to reach the photosensitive area. Metal trace shading, dielectric layer reflection, and absorption all contribute to significant photon loss. As pixel sizes shrink to sub-2 μm nodes, the ratio of the interconnect stack height to photodiode area continues to increase, drastically worsening this shading effect and causing FSI to reach its physical limits in terms of sensitivity and quantum efficiency.

In 2009, Sony made many back-side illuminated (BSI) structures for the first time. The design of the center is to dilute and turn the whole wafer, place the photodiode directly on the surface where the light enters on the back of the chip, and move the metal wiring layer below the place where the light feels. In this way, light can reach photodiode through any wiring structure, the aperture ratio is almost 100, and the highest quantum efficiency is over 95% [9]. The design of this "unstoppable optical path" also greatly improves the sensitivity of light entering from a large angle, and the design of the back focal length of the lens becomes more free.

However, BSI is still limited by the basic arrangement that photojournal and transistor are in the same plane in pixel. Stacked structure exceeds this limit by combining parts in three dimensions. Pixel layer and logic circuit layer are made separately, and then connected up and down by through-silicon vias or copper-copper direct bonding. In this way, the pixel layer can maximize the fill factor, and the logic layer can be put into more complex readout circuits and high-speed processing units. On this basis, Sony introduced a dual-layer transistor pixel structure. Photodiode, reset transistor, amplifier transistor and select transistor are divided into two silicon substrates. Photodiode uses the pixel position above to increase the saturated signal to twice that of the previous BSI, thus greatly expanding the dynamic range. Amplifier transistor is no longer limited by the size of the place where the light is felt, and can increase the size of the gate many times. This can directly reduce 1/f noise and random telegraph noise from the physical position of the part [3]. Sony LYTIA 901 sensor introduced in 2023 is loaded with this dual-layer transistor pixel structure, which indicates that the functional layering technology in pixels has entered the mass production stage [3].

3.2. Low-noise pixel structure research

The difference of noise is due to the working mechanism of shutter. "rolling shutter (RS) receives light line by line. There is no need for storage node in pixel, so the circuit is simple and the reading node is as low as 1.2 e^- However, if you shoot fast-moving things, the rolling shutter effect will appear. Global shutter (GS) exposes all pixels at the same time, so it can eliminate the distortion caused by motion. However, additional storage nodes are needed in each pixel. In the process of readout, the dark current of the storage node will continue to accumulate, so the noise floor will become high. At the same time, the structure of storage will occupy the sensitive place, so the fill factor and sensitivity will decrease. Therefore, the readout noise of GS is usually about twice that of RS.

In order to overcome this contradiction, the Sony Pregius S series combines back-illuminated photosensitive, stacked logic layers and global shutter. On the one hand, the size of the pixel is reduced to 2.74 μm, on the other hand, the readout noise is suppressed to 2.1 e^-, and the saturation capacity density of 1265 E/μ m is realized [10], which proves that the VB1940 series of noise ST Microelectronics uses different methods without great sacrifice when using GS. For the first time, you can automatically and intelligently switch between global shutter and rolling shutter. Use the accuracy of GS when moving quickly, and switch to RS when the light is dark, so that it can use the low noise and 18-bit HDR pipeline of RS, which destroys the common sense that the overall GS sensitivity of the previous system decreases [11].

4. REVIEW OF PIXEL STRUCTURE RESEARCH IN SPECIAL APPLICATION CASE

Because sensors are used more and more in automotive, detection and mobile applications, it is an important problem to expand dynamic range and reduce pixels. The brightness amplitude of natural scene is much larger than the traditional pixel linear range. However, the sub-micron objects meet the physical limit due to diffraction and full-well degradation.

4.1. High Dynamic Range (HDR) Pixel structure

To extend the dynamic range, there are three main technical paths for HDR pixel structure. The multi-exposure synthesis scheme directly uses standard pixels, and merges short, medium and long exposure frames off-chip or on-chip without changing the pixel structure. It has the best process compatibility, but the time offset between different frames will introduce motion artifacts. Staggered HDR and other improved timing schemes suppress ghosting by shortening the inter-line exposure interval, but cannot fundamentally eliminate the time difference between multiple frames. The logarithmic response scheme introduces a subthreshold operating MOS transistor in the pixel, so that the output signal is logarithmically related to the photocurrent. A single exposure can compress the maximum brightness range (more than 120 dB) and completely eliminate motion artifacts. However, the cost is low sensitivity in dark areas and serious fixed-mode noise caused by threshold mismatch between pixels, making the low-light image quality far inferior to that of linear response. The single-exposure intra-pixel scheme has become the current mainstream. Its common feature is that only one exposure is needed to complete HDR, fundamentally eliminating motion artifacts. Among them, the dual conversion gain (DCG) technology outputs high gain and low gain signals simultaneously in one exposure and fuses them outside the pixel. It has a simple structure and has been widely used in mobile sensors. The lateral overflow integral capacitor (LOFIC) technology adds a large capacity storage capacitor inside the pixel. When the photodiode overflows from the full well, the excess charge is collected by the capacitor, which greatly improves the equivalent full well capacity and the dynamic range can reach more than 110 dB. However, the capacitor occupies an extra pixel area [12]. The dual photodiode scheme realizes sensitivity splitting in the light domain through photosensitive areas of different areas. The current HDR pixel structure is moving from a single scheme to multi-mechanism collaboration, such as the combination of DCG and multiple exposures, and the fusion of LOFIC and 3D stacking, in order to achieve the comprehensive optimization of dynamic range, motion fidelity, and noise consistency.

5. SUMMARY AND FUTURE OUTLOOK

5.1 Current research findings and limitations

From the theoretical foundation laid by passive to 4T-APS, and through multiple rounds of innovation in back-illuminated, stacked, global shutter, and HDR structures, CMOS pixels have continued to advance in core performance indicators. Currently, the peak quantum efficiency of back-illuminated sCMOS has reached 95%, the readout noise of security-grade sensors has been reduced to 0.63 e^-, and the extreme mode of sCMOS has achieved a readout noise of 0.29 e^-; 3D stacked LOFIC technology has demonstrated a single-exposure dynamic range of 120 dB. However, compared to the theoretical limits, the gap remains significant: color imaging, due to the absorption filters, results in a system-level photon utilization efficiency of only about 33%, far below the theoretical upper limit of 100%; photon counting requires readout noise <0.15 e^-, currently about twice the level; room temperature dark current is about 26.8 e^-/s in global shutter pixels, while the target value required for low-light imaging is <6.25 e^-/s, a difference of more than 4 times; the dynamic range of a single exposure is still about 20 dB short of the practical physical limit of about 140 dB. Power consumption control and adaptability to extreme environments are also weak points; dark current increases exponentially at high temperatures, and while integrating AI into pixels can compress data bandwidth, it exacerbates the heat dissipation challenge. These gaps constitute clear targets for subsequent technological breakthroughs. A quantitative comparison of these gaps is provided in Table 1.

Table 1. Technical-Physical Limits Gap Comparison Table.

Index	Current level	Theoretical Limit/Target	Gap magnification
Quantum efficiency of color systems	~33%	100%	~3×
Readout noise	0.29 e^-	0.15 e^-	~2×
Dark current at room temperature	26.8 e^-/s	<6.25 e^-/s	>4×
Single exposure dynamic range	120 dB	~140 dB	~10×
Pixel integration density	~4-6 (2D) / 200+ (3D)	>1000 (3D)	~1 order of magnitude

5.2 Future Research Directions

The International Devices and Systems Roadmap (IRDS) 2024 proposes that semiconductor development in the next 15 years needs to be promoted in tandem with “More Moore” and “More than Moore”—the former improves integration density through new structures such as 3D stacking, while the latter achieves system value-added through heterogeneous

integration and functional fusion [13]. The future evolution of CIS pixel structure is unfolding along this dual-track framework. AI collaboration and in-pixel computing belong to the sensor-processor fusion path under the "More than Moore" framework. On-chip integrated DNN CIS has achieved target recognition in 0.8 μm pixels, and the in-pixel processing scheme without ADC proposes to directly generate neural network activation values at the pixel layer, eliminating the need for reading out the original data. IRDS predicts that in the next 10-15 years, sensors will move from "imaging first, then processing" to "perception as computing", and the output will change from pixel data to semantic features, with data bandwidth compressed by 1-2 orders of magnitude, directly serving the latency-sensitive needs of autonomous driving. Quantum dot enhancement and heterogeneous integration occupy the intersection of IRDS's "Beyond CMOS" and "More than Moore" paradigms. Quantum dots can be used to tailor absorption bands through size control, and have already achieved wide-spectrum monolithic CMOS imaging of 400-1300 nm with a fill factor approaching 100%. Photon counting and HDR breakthroughs represent a continuation of "More Moore" at the pixel-level. Double-layer transistor pixels, by increasing the amplification tube area to reduce 1/f noise, combined with related multiple sampling, are expected to break the 0.15 e^- readout noise threshold by 2028-2030, achieving room-temperature photon counting. Integrating the IRDS framework, the CIS pixel structure will evolve towards a general-purpose intelligent sensing front-end, based on three-dimensional heterogeneous integration and supported by in-pixel AI and photon counting. The corresponding device roadmap and key technology anchors for More Moore scaling are shown in Figure 2.

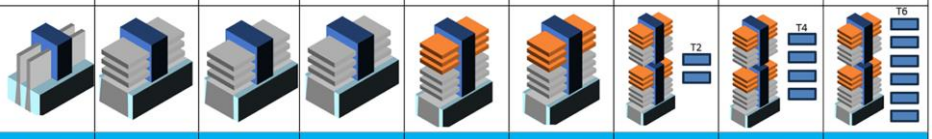
YEAR OF PRODUCTION	2024	2025	2027	2029	2031	2033	2035	2037	2039
	G48M24	G48M22	G48M22	G46M20	G44M18	G44M16	G42M14/T2	G42M14/T4	G42M14/T6
Logic industry "Node Range" Labeling	"3nm" Enhanced	"2nm"	"1.4nm"	"A10 eq"	"A7 eq"	"A5 eq"	"A3.5 eq"	"A2.5 eq"	"A1.8 eq"
Fine-pitch 3D integration scheme	Stacking	Stacking	Stacking	Stacking	CFET	CFET	3DVLSI	3DVLSI	3DVLSI
Logic device structure options	finFET	LGAA	LGAA	LGAA	CFET	CFET	CFET.3D	CFET.3D	CFET.3D
Backside structure options	-	Backside via	Direct contact	Decap+ESD	Active devices	Active devices	Active devices	Active devices	Active devices
Platform device for logic	finFET	LGAA	LGAA	LGAA	CFET	CFET	CFET	CFET	CFET
									
LOGIC TECHNOLOGY ANCHORS									
Device technology inflection	Taller fin	LGAA	LGAA	LGAA	CFET	CFET	3DVLSI	3DVLSI	3DVLSI
Patterning technology inflection for Mx interconnect	193i, EUV DP	193i, EUV DP	193i, High-NA EUV	193i, High-NA EUV	193i, High-NA EUV	193i, High-NA EUV	193i, High-NA EUV	193i, High-NA EUV	193i, High-NA EUV
Specialty transistors add-ons	-	-	-	2D, IGZO	2D, IGZO	2D, IGZO	2D, IGZO	2D, IGZO	2D, IGZO
Local interconnect inflection	Self-aligned vias	Backside via	Backside contact Substractive metal	Backside contact Substractive metal	Backside contact Active backside	Tier-to-tier via Active interconnect	Tier-to-tier via Active interconnect	Tier-to-tier via Active interconnect	Tier-to-tier via Active interconnect
Process technology inflection	Channel, RMG	Lateral/AtomicEtch	Low-temp contact	Low-temp contact	CFET	CFET	3DVLSI	3DVLSI	3DVLSI
Heterogenous stacking generation inflection	2.5D Chiplet, Mem-on-Logic	Backside Power	Backside Fabrics	Backside Fabrics	CFET	CFET	3DVLSI	3DVLSI	3DVLSI

Figure 2. Device Roadmap and Technology Anchors for More Moore Scaling.

REFERENCES

[1] Yu, F., Yusuke, T., Kamel, M., Keita, Y., Shoji, K.,Keiichiro, K. (2025). Programmable Dynamic Range Extension up to 110 dB Based on ChargeSplitting Method with 4-Tap CMOS Image Sensor. 2025 International Image Sensor Workshop, R05.2, https://imagesensors.org/Past%20Workshops/2025%20Workshop/2025%20Papers/R05.2_Feng.pdf

[2] ST Microelectronics. (2026). ST BrightSense a new complete portfolio of CMOS image sensors. https://www.st.com/content/st_com/en/campaigns/st-brightsense-imag-mccmos.html

[3] Sony. (2025). Sony Semiconductor Solutions to Release Approx. 200-Effective Megapixel Image Sensor for Mobile Applications with Built-in AI Technology Achieving high definition and high image quality for high-powered zooming on monocular cameras. Sony Semiconductor Solutions Group. https://www.sony-semicon.com/en/info/2025/2025112701.html

[4] Weckler, G.P. (1967). Operation of p-n junction photodetectors in a photon flux integrating mode," IEEE Journal of Solid-State Circuits, vol. 2, pp. 65-73.

[5] Fossum, E.R. (1993). Active pixel sensors: Are CCDs dinosaurs? Proc. SPIE, vol. 1900, pp. 2-14.

[6] Mendis, S., Kemeny, S.E., Fossum E.R. (1994). CMOS active pixel image sensor, IEEE Trans, Electron Devices, vol. 41, no. 3, pp. 452-453.

[7] Fossum, E.R. (1995). CMOS image sensors: Electronic camera-on-a-chip, IEDM Tech. Dig. pp. 17-25.

[8] Fossum, E.R. (1997). CMOS image sensors: Electronic camera-on-a-chip, IEEE Trans. Electron Devices, vol. 44, no. 10, pp. 1689-1698.

[9] Sony. (2010). Sony commercializes world's first*1 16.41 Megapixel "Exmor R™" back-illuminated CMOS image sensors for mobile phones. https://www.sony.com/en/SonyInfo/News/Press/201010/10-137E/

[10] Sony. (2024). Global Shutter Technology Pregius™ / Pregius S™. Sony Semiconductor Solutions Group. https://www.sony-semicon.com/en/technology/industry/pregius.html
[11] ST Microelectronics. (2026). VB1940 The automotive-grade 5.1 MP image sensor features global shutter and roll mode, ensuring full representation of image resolution and performance when selecting NIR and RGB. https://www.st.com.cn/zh/imaging-and-photonics-solutions/vb1940.html
[12] Omnivision. (2025). TheiaCel™ Technology A New Era for Single-Exposure HDR. https://www.ovt.com/technologies/theiacel-technology/
[13] IEEE IRDS. (2024). International Roadmap for Devices and Systems 2024 Edition. https://irds.ieee.org/images/files/pdf/2024/2024IRDS_MM.pdf