Review of Quanta Image Sensors for Ultralow-Light Imaging

Jiaju Ma, *Member, IEEE*, Stanley Chan, *Senior Member, IEEE*, and Eric R. Fossum, *Fellow, IEEE*

Abstract — **The quanta image sensor (QIS) is a photon-counting image sensor that has been implemented using different electron devices, including impact ionization-gain devices, such as the single-photon avalanche detectors (SPADs), and low-capacitance, high conversion-gain devices, such as modi ed CMOS image sensors (CIS) with deep subelectron read noise and/or low noise readout signal chains. This article primarily focuses on CIS QIS, but recent progress of both types is addressed. Signal processing progress, such as denoising, critical to improving apparent signal-to-noise ratio, is also reviewed as an enabling coinnovation.**

Index Terms — **CMOS image sensor (CIS), denoising, image quality, low-light sensor, photon-counting image sensor, quanta image sensor (QIS), subelectron read noise.**

I. Introduction

COUNTING every photon is as sensitive as physics presently allows in measuring light. To count photons incident on the faceplate, optical losses must be minimized, detector quantum and collection efficiencies must be maximized, and detector dead times minimized. Measurement of ultralow quanta (light) flux using single photomultiplier tube (PMT) detector photon counting was suggested as early as the 1960s, e.g., [1]–[3]. A digital photon-counting image sensor using APDs was suggested by Nippon Hōsō Kyōkai (NHK) [4]. In 1996, a hybridized photon-counting image sensor readout integrated circuit (ROIC) was investigated by Jet Propulsion Laboratory (JPL) [5] and the first solid-state single-photon avalanche detector (SPAD) was introduced [6]. In 2005, a new imaging paradigm based on photon counting

was described by Fossum [7] that considered a future pixel pitch of 0.5 μm or less and very limited full-well capacity (FWC). A similar concept was proposed again in 2009 by École polytechnique fédérale de Lausanne (EPFL) [8]. Such a device is now often referred to as a quanta image sensor (QIS) [9].

Various photon-counting image sensors were reported in a special issue of *Sensors* [10]. Most photon-counting image sensors are actually photoelectron-counting devices, with reflection and quantum efficiency (QE) loss, carrier collection loss, and detector dead time presumed to be acceptable, but not perfect. The detection of single electrons with deep sub-electron input-referred read noise (DSERN) has enabled the possibility of room-temperature megapixel photon-counting image sensors over the past ten years, with the assumption of high QE, or high photon-detection efficiency, which takes into account detector dead time. To achieve DSERN, two primary methods are used. The first is carrier-gain through the use of high electric field impact ionization either in avalanche diodes or through repeated high clock voltage charge transfer in an "impactron" [11] or electron multiplying (EM) charge-coupled device (CCD) [12]. The second method is the use of charge transfer devices such as a CCD or CMOS image sensor (CIS) with high conversion gain (CG) achieved through ultralow sense node capacitance and/or low noise readout electronics. The required read noise was suggested by Teranishi in 2011 to be less than 0.3e rms [13], [14] and later reduced to 0.15e rms in 2013 [15]. SPAD pixels typically achieve DSERN with ease. The first successful CIS-type pixel to achieve DSERN and demonstrate electron quantization was reported in 2015 [16], [17]. Each approach has advantages and disadvantages.

The purpose of this review article is to provide a useful overview and digest of progress in QIS realization, and pointers to the literature that has developed in this field. The article contains three major sections. First is a general discussion of the QIS and its imaging performance. QIS devices have been implemented using CIS-type principles and technology (referred to as CIS QIS) and SPAD devices (referred to as SPAD QIS). A brief review of CIS QIS and SPAD QIS devices will be presented along with thoughts on where each technology may be going.

Section II discusses the recent advances in ultralow noise imaging devices that can operate as CIS-QIS but which also retain legacy advantages of CIS devices. Such devices have benefitted from the technology developed for CIS QIS.

Photon-counting image sensors like the QIS are often operated in low quanta flux environments where photon shot noise
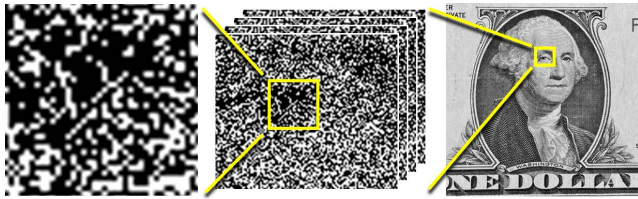
Fig. 1. QIS concept showing spatial distribution of binary jot outputs (left), expanded view of jot output bit planes at different time slices (center), and gray-scale image pixels (right) formed from spatio-temporal neighborhoods of jots.

limits the detection of signal-to-noise ratio (SNR) in the range $0 < SNR < 10$. Computational imaging approaches have been developed to improve apparent image quality through algorithmic and machine learning-based denoising, motion deblurring, and SNR enhancement of moving objects, and make these devices useful for machine vision and consumer use in low quanta flux regimes. Progress in this area is reviewed in Section III.

QIS devices will find applications where imaging in ultralow light is essential. These applications include security, night vision, space science, life sciences, biotech, quantum computing, aerospace, defense, and possibly automotive and consumer smartphones.

II. QIS CONCEPT

A. QIS Imaging Performance (Theoretical)

The QIS consists of an array of specialized pixels referred to as jots that are essentially binary in nature (indicating the arrival of at least one photoelectron, or not.) The QIS was originally envisioned to consist of millions or billions of small-pitch, low FWC jots readout at high frame rates, and thus very high bit rates. The concept originated when contemplating a future image sensor scaled to small pixel pitch and low FWC [7]. Image pixels are created from a local spatio-temporal ensemble of jot outputs (see Fig. 1) that are logically "zero" (no photoelectron) or "one" (at least one photoelectron). Bit density (D) is the number of logic "ones" divided by the total number of bits readout. It could be for a single jot readout many times (e.g., many frames) or a group of jots readout for one or more frames. The image sensor performance of QIS devices was analyzed by Fossum [15] for the expected value of D as a function of the average number of photons or photoelectrons that arrive at the jot during the exposure period, called the quanta exposure (H), the input-referred SNR_H, the dynamic range (DR), the bit error rate (BER) as a function of read noise, and other properties. In general, for $H \ll 1$, the performance is linear, but then approaching $H = 1$, the response becomes sublinear with a substantial overexposure latitude. This nonlinearity is fundamental and due to the statistical arrival of photons that are well described by the Poisson distribution probability mass function, which is the underlying cause of photon shot noise in image capture.

Plotting D-log H yields an S-shaped curve as illustrated in Fig. 2. The S-shaped D-log H curve has been known since 1890 [18] where, in this case, D is grain density in developed photographic plates, and H is the light exposure.
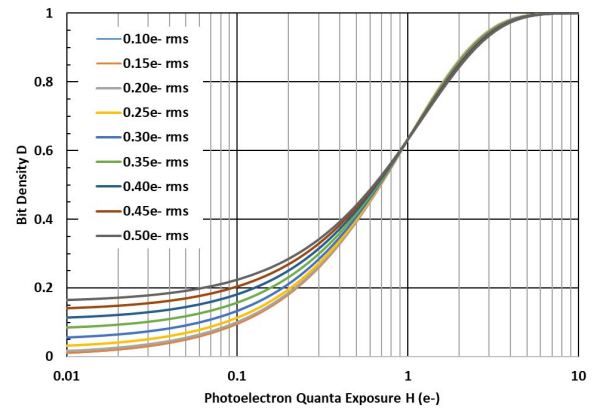

Fig. 2. Bit density (D) as a function of quanta exposure (H) calculated for a 1bQIS for different input-referred read noise levels. Adapted from [20].

It was observed in a time before the quantum of light, the photon, was described by Planck and Einstein in the early 1900s. In fact, the same basic Poisson statistics are behind the D-log H characteristics of Hurter and Driffield, and those of the QIS.

The bit density, noise and SNR predicted by the 2013 QIS model was first experimentally verified using a SPAD QIS in 2015 [19]. Measurement of the D–H characteristic can be used to estimate read noise and quantizer thresholds in CIS QIS devices [20], [21].

The binary QIS concept was expanded to include low bit-depth output—i.e., effective FWC greater than unity. The binary QIS is now referred to as a 1bQIS and the latter as a multibit QIS, or mbQIS. In the mbQIS, the low bit-depth digital value is equal to the number of electrons readout. Multibit quantizers can be programmable to trade power and read out speed with bit depth and concomitant nonlinearity, e.g., [16], [22], [23], [24]. This 1–7b photon number resolution capability differentiates mbQIS from higher read noise and higher bit resolution ($\sim$10–14b) regular CIS devices. However, if anything, this differentiation has become blurred as regular CIS devices have emerged with DSERN, as described in Section II-B. Photon-counting error rates in 1bQIS and mbQIS were analyzed in 2016 [25].

It is noted that while the QIS is a binary-output image sensor, it differs from some binary sensors that have appeared in the literature over the years, wherein the threshold for triggering a change in output value typically represents a few or perhaps many photons, e.g., [8], [26].

B. Implementation: CIS QIS and SPAD QIS

In principle, any device that can detect photoelectrons with less than 0.15–0.30e$^-$ rms read noise to achieve low BER (i.e., BER < 0.0005–0.005 bit-errors/read) can be used as a QIS device. For example, a cooled EMCCD [12] can operate as a 1bQIS, albeit with a slower readout rate (but not so well as a mbQIS due to gain noise), and a cooled CCD with "skipper readout" (many nondestructive reads of a pixel) can also be used as 1bQIS or mbQIS, albeit with an even lower frame rate [27].

Two major approaches seem promising at this time for room temperature (RT) application. The first is CMOS image

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

MA *et al.*: REVIEW OF QISs FOR ULTRALOW-LIGHT IMAGING 3

TABLE I
EXAMPLES OF THE REPORTED QUANTA IMAGE SENSOR (QIS) DEVICES AND THEIR CHARACTERISTICS. (BSI = BACKSIDE ILLUMINATION, LV = LOW VOLTAGE, FPS = FRAMES PER SECOND, RT DCR = ROOM TEMPERATURE DARK COUNT RATE, QE = QUANTUM EFFICIENCY, AND PDE = PHOTON DETECTION EFFICIENCY)

Year	Who	Type	3D BSI	LV	Depth 1b/mb	Pitch (µm)	Res. (Mpix)	FPS (fps)	Power (mW)	RT DCR (e-/s)	QE/ PDE	Ref
2014	Edin./ST	SPAD			1b	8.0	0.077	5,000	69	312	-	[33]
2014	EPFL	SPAD			1b	24.0	0.065	156,000	1650	350	14%	[35]
2015	MIT/LL	SPAD			1b/mb-7	25.0	0.065	8,000	-	>1000	-	[36]
2015	Dartmouth	CIS		✓	Analog	1.1	0.001	n/a	n/a	<1	-	[17]
2016	Edin.ST	SPAD	✓		mb-12	7.83	0.015	500	70	<200	12%	[37]
2017	Dartmouth	CIS	✓	✓	1b	1.1	1.0x20	1,000	19	<1	80%	[28]
2019	EPFL	SPAD			1b	16.4	0.262	97,700	700	7.5	-	[38]
2019	Edin./ST/HWU	SPAD	✓		mb-14	9.2	0.065	30	78	20	23%	[39]
2019	Panasonic	VAPD			1b	6.0	0.160	60	-	100	-	[40]
2020	Canon/EPFL	SPAD			1b	9.4	0.5x2	24,000	1070	2	3.6%	[41]
2021	Canon	SPAD	✓		mb-11	6.4	3.2	60	-	1.8	69%	[42]
2021	Sony	SPAD-	✓		mb-9	12.2	0.042	60	-	35	62%	[43]
2021	FBK	SPAD			1b	7.0	0.0002	-	-	>1000	-	[44]
2021	Gigajot	CIS	✓	✓	mb-12	2.2	4.194	60	550	0.2	84%	[31]
2021	Gigajot	CIS	✓	✓	mb-12	1.1	16.777	30	600	0.02	80%	[32]
2022	Gigajot/Dart	CIS	✓	✓	1b	1.1	2.097	500	68	<1	80%	[21]

sensor-based QIS (CIS QIS) developed at Dartmouth since 2011, and the second is a SPAD-based jot device (SPAD QIS). The selected example devices from the literature are presented in Table I.

1) CIS QIS: The CIS QIS approach requires a pixel with high CG and/or low input-referred read noise, and a quantizer circuit to convert the analog-sensed voltage signal to a digital value (one or more bits in depth, corresponding to the electron number). The first 1 kpix CIS QIS was reported in 2015 [17]. A 1 Mpix 3D-stacked-backside illumination (BSI)-CIS QIS was reported in 2017 [28] with 1.1 µm pixel pitch, 1 kfps frame rate, 17.6 mW power dissipation, 0.21e rms avg read noise, and 0.2e /s dark count rate. In fact, 20 different 1 Mpix QIS devices with varying designs were integrated on a single chip so this might be considered as a 20 Mpix QIS.

The advantages of the CIS QIS approach are small pixels (e.g., 1 µm pitch), high resolution (e.g., >100 Mpixels), very high photon detection efficiency (PDE), relatively low power, low electric field strengths, low DCR, photon number resolution (multibit QIS), and likely high manufacturing yield and lower cost for a given resolution. An indirect advantage is leverage from the advancement of regular CIS pixel technology and shrink, requiring less unique detector device engineering from generation to generation.

Drawbacks to the CIS QIS are primarily in control of the quantizer threshold voltage(s) across the sensor. Reduction in read noise and/or increased CG will ameliorate this drawback, as would self-calibration. Several techniques have been developed to characterize read noise and quantizer threshold [20], [29], [30].

QIS technology is being applied to achieve DSERN performance in CIS devices and enable ultralow-light image capture capability along with high-DR (HDR) and other features found in commercial and consumer CIS devices [31], [32].

2) SPAD QIS: The SPAD QIS, used to first verify QIS imaging performance predictions, has made strong progress recently. In 2014, a 77 kpix SPAD QIS was reported by the Edinburgh and STMicroelectronics (ST) Micro [19], [33], [34] and a 65 kpix SPAD QIS was published by EPFL [35] with 8 and 24 µm pixel pitches, and 5.14 and 156 kfps frame rates, respectively. In 2015, Massachusetts Institute of Technology (MIT) Lincoln Labs reported a 65 kpix SPAD QIS with 25 µm pixel pitch and 8 kfps frame rate [36]. The first BSI-stacked mb-QIS with 7.83 µm pitch and 15 kpixels was reported in 2016 by Edinburgh and ST Micro [37].

By 2019, a 1/4 Mpix SPAD QIS was reported [38] as well as an improved 3-D BSI-stacked SPAD QIS [39]. A variation in a SPAD QIS (160 kpix) was presented by Panasonic using a vertical avalanche photodiode [40].

In 2020, the first 1 Mpixel SPAD QIS was reported (actually 2 × 0.5 Mpixel arrays) by a Canon/EPFL collaboration [41]. The SPAD QIS had a 9.4 µm pixel pitch with a 24 kfps frame rate with power dissipation of up to 535 mW for 0.5 Mpixel readout. Canon further progressed the technology to achieve 3.2 Mpix with a 6.39 µm pixel pitch and a 60 fps frame rate with DCR and PDE approaching CIS QIS levels using a 3-D-stacked BSI process. This mbQIS has an 11b pixel-parallel digital counter in the bottom tier to allow photon number resolution and HDR. Power dissipation was not reported [42]. A SPAD QIS with a pixel-parallel digital counter, (42.2 kpixels, 12.24 µm pixel pitch, and 60–250 fps) was reported by Sony at about the same time [43]. A novel 1-

T SPAD QIS test array (200 pixels, 7 µm pitch) with a single access transistor to the pixel was presented by Fondazione Bruno Kessler (FBK) [44].

The primary advantage of SPAD QIS results from the nearly instantaneous and large carrier-gain provided by the avalanche photodiode breakdown that is triggered by a photoelectron. The voltage pulse it creates can be used to time-stamp photon arrival permitting time-of-flight measurement. The gain can be turned "off" to provide a gating function. Once triggered, the avalanche feedback process results in no apparent read noise. The lack of read noise is usually balanced by lower PDE which relates to photoelectrons triggering the avalanche feedback process, and thus sometimes photoelectrons become lost and uncounted.

The dual-mode operability of SPAD QIS to gate and record photon arrival times, as well as provided QIS-mode imaging, is a strong potential advantage of SPAD QIS compared to present-day CIS QIS but can result in a larger pixel pitch.

The use of high internal electric fields needed to trigger avalanche and high gain is a weakness of SPADs, resulting in the need to isolate pixels, in turn leading to larger pixel pitches. The higher electric fields can exacerbate DCRs and potentially impact device yield. Die cost is a function of pixel size, resolution, and yield, so at the current time, SPAD QIS is expected to be more costly to manufacture than CIS QIS.

Power dissipation at higher photon count rates can cause large CV^2f power dissipation in the SPAD array (e.g., 1–10 W), which can exceed that of the readout circuits, due to high bias voltages and avalanche currents [45] that must recharge the full pixel capacitance with each photon arrival.

While the digital readout layer shrink will track digital circuit technology node improvement, pixel shrink at the SPAD layer may be more difficult to achieve and there may be little leverage from regular CIS technology improvements in terms of shrink aside from 3-D BSI stacking. However, earlier work in nano-sized APDs in 2007 may guide future SPAD shrink [46] and the minimum SPAD pixel size reported so far is 3 µm [47]. Scaling laws for SPADs were suggested in 2021 [48].

III. ACHIEVING DEEP-SUBELECTRON READ NOISE

In recent years, a significant amount of research effort has been spent on the reduction of read noise, for the development of QIS and the improvement of low-light imaging performance in CIS. Although there are a variety of approaches being explored for reducing the read noise, they can be summarized into two main categories, improving the CG of the pixel and reducing the voltage temporal noise of the in-pixel source follower (SF).

The improvement of pixel CG was realized in two ways: 1) reducing the floating diffusion (FD) capacitance and 2) replacing the in-pixel SF with high-gain amplifiers. Additionally, the reduction of the pixel SF temporal noise was demonstrated with buried-channel SFs and pMOS-based SFs. The correlated multiple sampling (CMS) is commonly used with other techniques to further lower the read noise.

The advancement of the CMOS manufacturing process also contributes to the reduction of read noise. The subelectron read noise performance was reported in [49] and [50] with standard
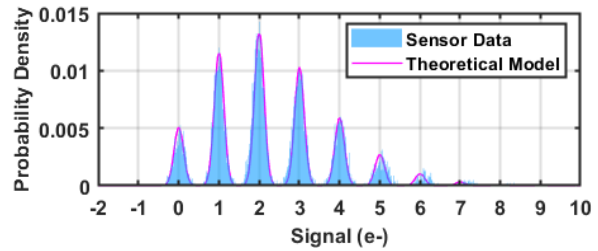

Fig. 3. Single-pixel PCH with 0.12e⁻ rms read noise measured at RT, reported in [32].
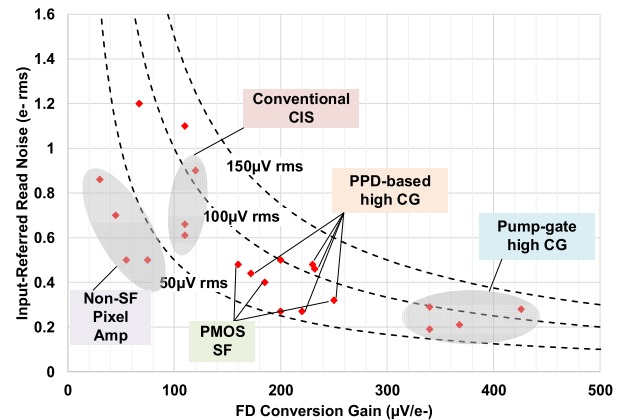

Fig. 4. Read noise and FD CG performance of the selective recent CIS and QIS. The dashed reference curves show the input-referred read noise in voltage (µV rms).

CIS devices fabricated in a 45 nm standard CIS process and a typical pixel CG of 110–120 µV/e⁻. The voltage read noise of these devices is reduced to about 100 µV rms without CMS and 70 µV rms with CMS.

The read noise performance of the recently published low-noise CIS is summarized in Table II. Among these listed results, the lowest input-referred read noise was reported in [32] by Ma. Read noise of 0.19e⁻ rms was achieved in a 16.7 Mpix CIS QIS with 1.1 µm pixels. This record-low read noise was realized with a high CG of 340 µV/e⁻, enabled by the pump-gate pixel structure. As shown in Fig. 3, a photon-counting histogram (PCH) with 0.12e⁻ rms read noise is reported in this work. The discrete photo-electron peaks in the histogram are well aligned with the Poisson–Gaussian model, which demonstrates the reliable photon-counting capability of the sensor. A scatter plot of the read noise of these sensors vs. FD CG is shown in Fig. 4. The dashed reference curves show the input-referred read noise in voltage (µV rms). Without considering the difference of the FD CG, the lowest voltage read noise (∼25 µV rms) was reported in Ge [51] and Lotto [52]. The reduction of voltage read noise was realized with in-pixel non-SF amplifiers with a significantly higher voltage gain. Subelectron read noise was also demonstrated with pMOS-based SF and buried-channel SF [53]–[57]. Both devices demonstrated effective noise reduction compared to the conventional nMOS-based surface-channel SF: ∼80 µV rms voltage read noise (pMOS) without CMS and 45 µV rms voltage read noise (buried-channel nMOS) with CMS.

These read noise reduction techniques are discussed in more detail in the sections below.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

MA *et al.*: REVIEW OF QISs FOR ULTRALOW-LIGHT IMAGING
5

TABLE II
SUMMARY OF THE READ NOISE PERFORMANCE
OF THE SELECTIVE RECENT CIS AND QIS

Technique	Read Noise (e- rms)	FD CG (uV/e-)	Voltage RN (µV rms)	Pixel Size (µm)	CMS	Process (nm)	Ref
PPD-based high CG	0.78	NA	NA	2.9	NA	NA	[107]
PPD-based high CG	0.27	220	59.4	11.2	Yes	110	[58]
PPD-based high CG	0.44	172	75.68	NA	128	110	[59]
PPD-based high CG	0.46	232	106.72	5.5	No	180	[60]
PMOS SF	0.48	160	76.8	6.5	No	180	[53]
PMOS SF	0.4	185	74	7.5	No	180	[54]
PMOS SF	0.32	250	80	10	4	180	[55]
Buried-channel SF	0.7	45	31.5	NA	4	180	[57]
Non-SF pixel amp	0.42	NA	NA	10	Yes	180	[69]
Non-SF pixel amp	0.5	55	27.5	11	No	180	[51]
Non-SF pixel amp	0.5	75	37.5	1.45	2	90/55	[70]
Non-SF pixel amp	0.86	30	25.8	11	No	180	[52]
Conventional PPD	0.66	110	72.6	1.1	5	45/65	[49]
Conventional PPD	0.9	120	108	0.9	No	45/65	[50]
Conventional PPD	0.61	110	67.1	7.1	NA	110	[62]
Conventional PPD	1.1	110	121	1.4	16	90	[106]
Pump-gate	0.19	340	64.6	1.1	8	45/65	[32]
	0.29	340	98.6		No		
Pump-gate	0.27	200	54	2.2	16	45/65	[31]
	0.5	200	100		No		
Pump-gate	0.21	368	77.28	1.1	16	45/65	[28]
Pump-gate	0.28	426	119.28	1.4	8	65	[17]
Pump-gate	0.48	230	110.4	1	8	65	[61]

Fig. 5. Pump-gate pixel structure reported in [64].

A. Small FD Capacitance

High pixel CG is demonstrated in multiple works with significantly reduced FD capacitance [17], [28], [31], [32], [55], [58]–[62]. The capacitance of the FD node in a standard CIS pixel includes a few components: 1) FD p-n junction capacitance; 2) FD to transfer gate (TG) overlap capacitance; 3) FD to reset gate (RG) overlap capacitance; 4) SF gate capacitance; and 5) intermetal capacitance. As the fabrication process advances, the gate oxide becomes thinner and the capacitance components 2)–4) increase proportionally. In the pixels with shared readout architecture [63], the FD node is coupled to multiple TGs, which proportionally increases the FD-TG overlap capacitance.

The FD total capacitance can be lowered by reducing one or multiple of these capacitance components. A pump-gate pixel structure was first reported [64] by Ma for the reduction or elimination of the FD-TG overlap capacitance with a distal FD. As shown in Fig. 5, a three-step electrostatic potential profile including a virtual-phase region is created in the pump-gate device to enable a complete charge transfer from the storage well (SW) to the distal FD node. This device was first fabricated [17] and 426 µV/e⁻ CG was demonstrated in 1.4 µm pixels, which is equivalent to a total FD capacitance of only 0.38 fF. In this work, DSERN (0.28e⁻ rms) was realized for the first time with CIS pixels due to the high CG and its
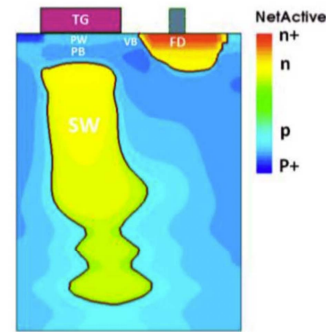
PCH demonstrated photon-counting capability. The pump-gate device was further improved [28], [31], [32] and recently implemented in commercial QIS products [65]. Despite the ultrasmall FD capacitance, good interpixel uniformity and low photon-response nonuniformity (PRNU) (1%) are realized in multimega-pixel HDR QIS devices [32].

New pixel structures were also introduced to reduce other FD capacitance components. In [28], [58], [59], and [66], the reset transistor was replaced with a gateless reset diode, often termed "punchthrough reset (PTR)," to eliminate the FD-RG overlap capacitance. With the PTR diode, the FD node is reset by increasing the positive bias voltage of the reset drain (RD) node. As shown in Fig. 6, the higher bias increases the depletion width surrounding the RD node and lowers the potential barrier between the FD-RD junction, which allows the electron current to flow from the FD to the RD. With the PTR, a higher supply voltage is needed to achieve an equivalently high FD reset voltage to preserve the FD voltage swing and the DR. This requires an additional positive charge pump or other on-chip high-voltage generators and increases the complexity of the sensor. Hence, a bootstrapping operation was introduced in [59] to increase the FD reset voltage in the PTR by manipulating the FD capacitance before and after the reset operation, without increasing the bias voltage on the RD node.

The improvement of CG was also reported in the standard CIS pixels with mild implant modifications. In [60], optimized n⁺ and lightly doped drain (LDD) implantation conditions were applied to the FD and the SF drain with lowered dose/energy to reduce the FD junction and the SF gate capacitance. A CG of 240 µV/e⁻ was demonstrated with these modifications, which is equivalent to 0.67 fF FD capacitance.

Novel SF devices are also explored to reduce the SF gate capacitance. A JFET-based pixel SF was proposed in [67]. This is a p-channel JFET SF created in the pixels with implantations. The FD node functions as both the sense node and the gate of the JFET. The JFET is biased with a constant current source, and the output voltage follows the FD voltage when the JFET is biased in the saturation region. The characterization results of this device are reported in [68], and an extremely high CG of 540 µV/e⁻ was measured from some pixels, which is equivalent to a FD capacitance of only 0.3 fF. However, a large across-device variation was also observed, likely due

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON ELECTRON DEVICES
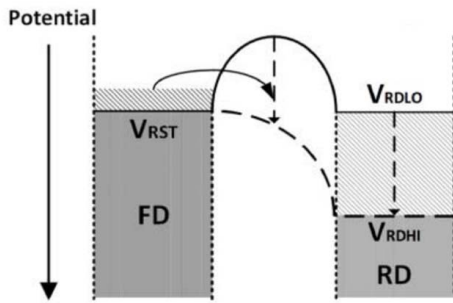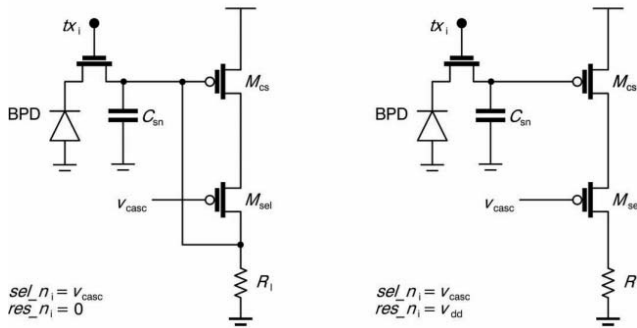

Fig. 6. Gateless reset diode reported in [28].

Fig. 7. Pixel-level common-source amplifier with a negative feedback and self-biased reset method, in reset configuration (left) and amplification configuration (right), reported in [52].

to the nonuniformity of the doping concentration of the JFET across the pixel array.

B. Non-SF High CG Pixels

Another interesting approach to enable high CG in CIS-based pixels is to replace the pixel SF with other amplifiers with a higher voltage gain. In [52], the pixel SF is replaced with a pixel-level common-source amplifier with column-wise load resistors. A nominal voltage gain of 10 V/V and 300 μV/e CG on the column output node were realized with this open-loop configuration. This yields a relatively low FD-referred CG of 30 μV/e. The correlated double sampling (CDS) operation was used to cancel the pixel-to-pixel variations of the amplifier offset induced by the mismatch of the threshold voltage of the common-source transistors. A self-biased reset method with negative feedback (Fig. 7) was used to compensate for the variations of the pixels' linear output swing. A 2.5% PRNU was realized with these compensation schemes, which is still higher than the typical performance of SF-based CIS pixels but remarkably low for pixels with open-loop amplifiers. The sensor achieved 0.86e rms read noise. Considering the relatively low CG on the FD node, the input-referred voltage noise achieved with this approach is as low as 25.8 μV rms, which is significantly lower than the voltage noise of the SF-based pixels.

A similar pixel-level voltage amplification architecture was also reported in [51] and [69] with an additional column-level
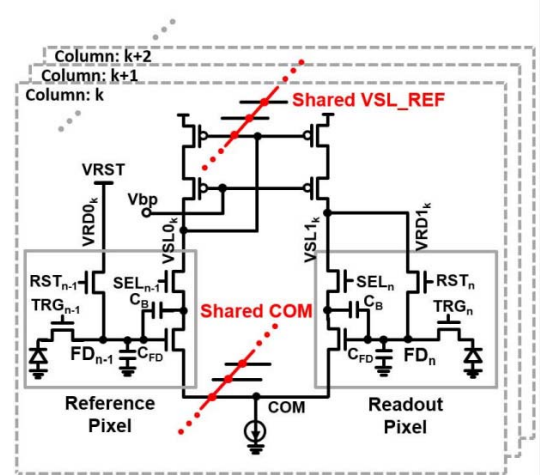

Fig. 8. In-pixel differential common-source amplifier, reported in [70].

sinc-type low-pass filter to further reduce the voltage noise. A minimum read noise of 0.31e rms and peak read noise of 0.42e rms were reported. However, the sensors suffer from large pixel-to-pixel CG variations (e.g., 240–2200 μV/e in [69]), which may limit the implementation of this technique in the applications that have strict requirements for PRNU.

With a slightly different approach, an in-pixel differential common-source amplifier was proposed in [70]. As shown in Fig. 8, the differential common-source amplifier is formed with a readout pixel and a reference pixel, providing a nominal voltage gain of about 7.5 V/V and a column-referred CG of 560 μV/e. The reference nodes, COM and VSL_REF, are connected in parallel among thousands of pixels that are simultaneously readout, which significantly increase the transistor size and reduce the temporal noise from the biasing transistors. This work realized 0.50e rms read noise and an improved PRNU of 2.5% compared to the single-ended configuration used in [51] and [69], which suggests better uniformity of the CG across the pixels.

C. SF Temporal Noise

In the SF-based CIS pixels, the temporal noise from the SF is usually the dominating noise source. The temporal noise in an SF device consists of thermal noise, 1/f noise, and random telegraph noise (RTN). Thermal noise is present in all electrical circuits, and its cause is well understood to be the thermal fluctuation of the charge carriers inside the electrical conductor [71]. Similarly, 1/f noise is present in almost all the electrical circuits. Its root cause, although has been extensively studied, is still largely debatable [72]–[80]. The popular theories include the fluctuation of the number of charge carriers in the transistor channel and the fluctuation of the mobility of the charge carriers. However, none of the models managed to explain all the experimental results. RTN is often present in a small portion of a large pixel array. The percentage of the RTN pixels can be lower than 100 ppm in a modern CIS. However, because of its high noise magnitude and trimodal noise signature, the RTN pixels are

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

MA *et al.*: REVIEW OF QISs FOR ULTRALOW-LIGHT IMAGING   7

usually shown in the low-light images as "blinking" pixels and have strong degradation to the image quality. The RTN in CIS is well known to be linked to the trapping/emission events of the defects-induced energy states inside the pixels, especially inside the Si–gate oxide interface in the SF channel, e.g., [81]–[93]. Other RTN sources have also been observed in CIS [83], [84], [93], such as the photodiode dark current induced RTN and the gate-induced drain leakage (GIDL)-induced RTN.

The use of a "buried channel" was first introduced in buried-channel charge-coupled devices (BCCDs) to reduce the interaction between the charge carrier and interface traps, thus improving charge transfer efficiency [94]. This concept was later expanded to the in-pixel SF devices to reduce the RTN and 1/f noise [56], [57], [85], [95]. The buried-channel SF (BSF) reported in [95] consists of a thin n-type channel located near the Si–SiO$_2$ interface and between the n$^+$ doped source and drain. Because of the n-type buried-channel doping, this device has a negative threshold voltage. When the device is biased in the saturation region, the negative voltage across the gate and the channel creates a potential barrier near the Si–SiO$_2$ interface with a barrier height more than several kT/q, which protects the charge carriers in the channel from the interface traps. In [95], a 50% read noise reduction compared to the surface-channel SFs with the same size and 205 μV rms input-referred read noise were reported. The effective noise reduction from the BSF was confirmed in [85], in which a 5× noise reduction at the 99.99% percentile and a 90× reduction of the RTN quantity compared to the surface-channel SFs were reported.

Additionally, reduction in 1/f noise and RTN was demonstrated with pMOS SF in multiple works [53]–[55], [96]–[99]. The lower noise of pMOS can be explained by the lower active trap density in pMOS because of the 10–20 times heavier effective masses of a hole in the oxide than that of an electron and a higher potential barrier for a hole to tunnel into SiO$_2$ [75], [100]. The pMOS SF can be implemented in CIS pixels with a hole-based p-type process [97]–[99], or more commonly in the modern CIS, with an in-pixel n-well made with implantations to host the pMOS SF [53]–[55], [101]. However, the n-well will inevitably increase the pixel size and reduce the fill factor. In [53], a thin-oxide pMOS SF was implemented and 0.48e rms input-referred read noise was realized, which is equivalent to 76.8 μV rms read noise in the voltage domain. This work was expanded in [55], and the input-referred read noise was further improved to 0.32e rms with 250 μV/e CG and CMS readout. In addition, in the pMOS SF reported in [101], a bulk-to-source connection was made to compensate for the body effect and improve the voltage gain of the SF.

As both 1/f noise and RTN are known to be inversely proportional to the gate size of the SF [79], [80], [91], [96], a larger SF size is desirable for the reduction of SF temporal noise. However, a larger SF also increases the capacitance on the FD node and reduces the CG. This tradeoff is discussed in [28] and [102]. Recently, a multigate SF was introduced as a possible solution to overcome this tradeoff with promising preliminary results [103].
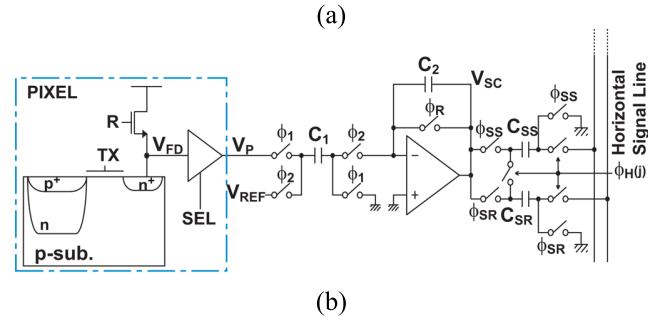

Fig. 9. Example implementation of CMS operation in (a) digital domain [114] and (b) analog domain [112].

D. CMS and Noise Filtering

The CDS readout is commonly used to in modern CIS to eliminate the FD reset kTC noise and reduce the SF thermal noise and 1/f noise [104]. As an expansion of CDS, CMS readout is often used to further reduce the read noise [17], [28], [31], [32], [49], [55], [57]–[59], [61], [70], [105]–[115]. With CMS, the pixel reset and signal voltage levels are sampled multiple times and the averages are subtracted. Hence, the pixel reset noise can be canceled through subtraction, just like CDS, and the thermal noise and 1/f noise can be further reduced with averaging. The CMS readout has been implemented in CIS in both digital and analog domains. Examples of the digital and analog implementation are shown in Fig. 9.

Compared to analog CMS, digital CMS requires a larger number of analog-to-digital converter (ADC) conversions, which results in a reduced frame rate and increased power consumption. The analog implementation is more time and power efficient; however, it is usually less efficient in noise reduction because of the additional kTC noise in the sample-and-hold circuitry. Novel circuit architectures are actively explored to overcome this tradeoff. For example, in [49] and [108], a selective digital CMS method was used to shorten the ADC conversion time needed for the multiple sampling. With this architecture, the pixel output is sampled simultaneously by a full-range ramp for large signal under strong illumination and a multiple sampling short ramp for small signal under dark conditions. This approach reduces the readout time needed for digital CMS while preserving the noise reduction efficiency, but it introduces additional complexity to the per-column ADC and the signal processing, as well as the chip area and power consumption.

The theoretical read reduction from CMS is as follows: $\sigma_{CMS} = \sigma_{CDS}/\sqrt{N}$, where σ_{CMS} and σ_{CDS} are the read noise with CMS and CDS, respectively; and N is the number of CMS cycles. However, the noise reduction observed in the experimental results often show lower

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE TRANSACTIONS ON ELECTRON DEVICES
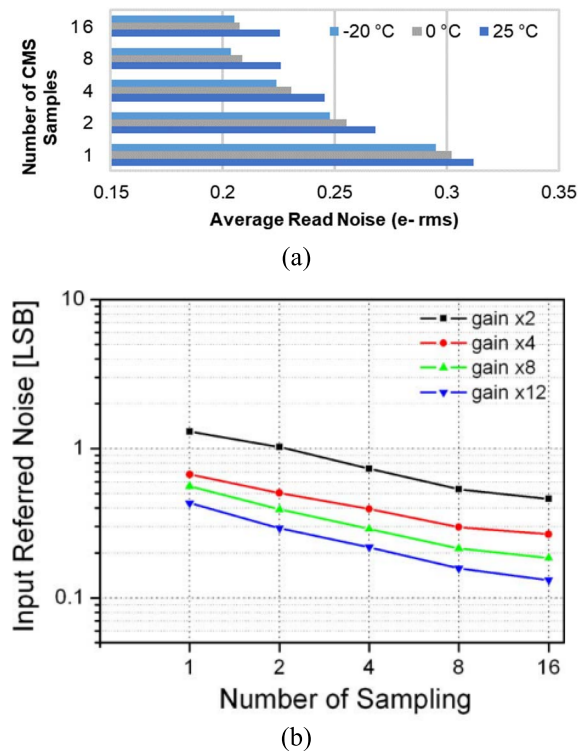

(a)

(b)

Fig. 10. Measured read noise versus number of CMS cycles (a) from [32] and (b) from [114].

efficiency than the theoretical model, especially with a large N (Fig. 10) [31], [32], [112], [114]. This phenomenon can be explained by lower frequency 1/f noise and the accumulation of the dark current on the FD node as the sampling time increases. As discussed in [115], a skipper-type of CMS operation will be the most efficient for the read noise reduction [116]–[118], as the effective sampling time can be kept short for each pair of the reset and signal samples to cancel the low-frequency noise and the accumulation of FD dark current. However, this technique requires a floating gate or similar types of readout architecture in the pixels, which reduces the CG on the FD node and increases the complexity of the pixel structure.

The reduction of read noise has also been demonstrated with other noise filtering methods by limiting the noise bandwidth of the readout circuit. A faster CDS operation with a shorter t between the two samples can effectively reduce the read noise [88], [119], and a similar reduction can be realized with a lower bias current of the pixel SF. However, both techniques have limitations with high-speed operation under high-light conditions when a large signal swing and fast settling time are needed.

E. Superior Low-Light Imaging With DSERN

Reducing read noise from 1e rms to DSERN levels brings somewhat surprising improvements to the ultralow-light imaging performance with CIS-based multibit QIS. As shown in Fig. 11, a CIS QIS sensor is compared with two industry-leading CISs for security and cellphone applications under ultralow-light conditions (10 and 128 mlux) with the same

exposure time and lens configurations. Despite the significantly smaller pixel size, the QIS provides remarkably better SNR and image quality, due to the ultralow read noise.

IV. SIGNAL PROCESSING FOR QIS

Data captured by a QIS is a three-dimensional space-time volume where each entry is a 1-bit or multibit digital number. Since in principle the jot size can be small and the temporal response can be fast, the binary outputs produced by the jots can be seen as repeated but independent measurements of the incident photon flux. A schematic of this image formation process is shown in Fig. 12. The process is a combination of color selection, photon arrival, noise injection, and quantization, among other sensor level modeling.

At the very basic level, the mathematical model of the measured jot value Y can be described by the following equation:

$$Y = \text{ADC}\ \text{CFA}\ \text{Poisson}(H + H_{\text{dark}}) + \text{Gauss}\left(0,\ ^2\right)$$

where H is the quanta exposure, H_{dark} is the dark current, and is the read noise standard deviation. The sum of the Poisson random variable and the additive Gaussian random variable accounts for the photon arrivals and the read noise, respectively.
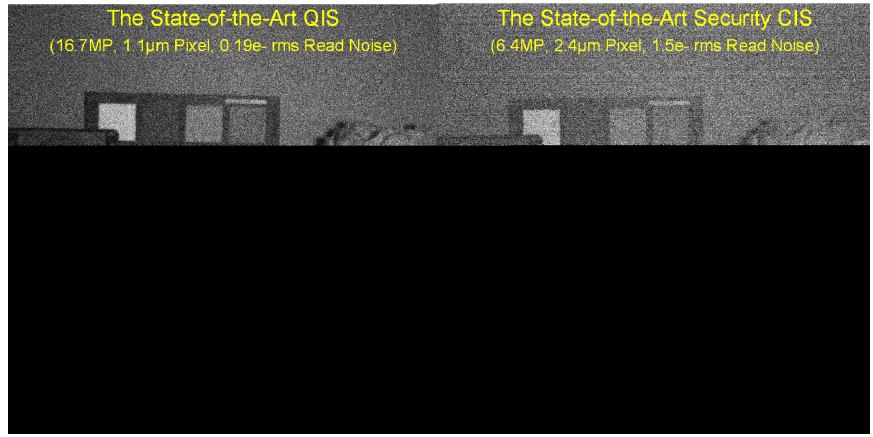
A color filter array (CFA) is applied to the measurement to give color, and an ADC is used to convert the voltage to digital bits. Assuming that the underlying exposure H does not change rapidly over space and time, the random variable Y is sampled repeatedly to produce the observed data.

Vetterli and colleagues at EPFL [8], [121], [122] had a precise abstraction of QIS, referring to it as an *oversampling* device because the information is embedded in the densely sampled measurements. The nonlinearity of the image formation makes the statistical properties of the data less straightforward compared to CIS [15], [123]–[125], and thus the signal extraction from the raw data to an actual image poses new challenges.

The rest of this section will describe the signal processing aspects of QIS. The mathematical model presented here is one level above the device modeling. What this means is that the model is applicable whenever the image formation follows a Poisson–Gaussian distribution, subject to different parameters, e.g., CIS QIS has a lower dark current than that of SPAD QIS. Because of the identical mathematical formulation, the algorithms are valid for both CIS QIS and SPAD QIS. In fact, the reported algorithms seldom distinguish themselves based on the particular technology [142] and [150].

A. Estimation for 1-Bit and Multibit QIS Signals

The basic building block of QIS signal processing is to consider Poisson (H) by ignoring the dark current and read noise. The ADC (or simply a threshold mechanism) will turn the measured voltage into a quantized random variable Y depending on the bit depth. For 1-bit signals, Y is binary with two states $Y = 1$ and $Y = 0$. The probability distribution of Y is $P[Y = 1] = 1\ e^{-H}$ and $P[Y = 0] = e^{-H}$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

MA *et al.*: REVIEW OF QISs FOR ULTRALOW-LIGHT IMAGING

9

(a)

(b)

Fig. 11. Low-light imaging comparison between the state-of-the-art multibit QIS (Gigajot GJ01611) and CIS. (a) Comparison with a security CIS that has a 4.76× larger pixel size under 10 mlux with 40 ms exposure time and F/1.4 lens. (b) Comparison with a cellphone CIS that has a 1.78× larger pixel size under 128 mlux with a 44 ms exposure time and F/1.6 lens. Images from the QIS are raw without advanced image enhancement such as denoising.
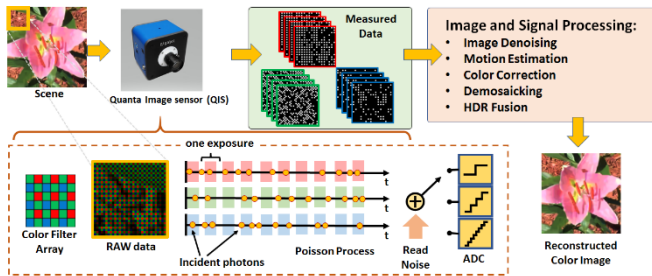

Fig. 12. Schematic illustration of the image formation of QIS. The incident flux is sampled rapidly using a binary (or a few bit) measurement. The goal of signal and image processing is to recover the underlying scene. Image courtesy: [120].

For multibit signals, it can be shown that if the saturation level is L, then [125]

$$P[Y = k] = \frac{H^k}{k!}e^{-H}, \quad \text{for } k = 0, 1, 2, \ldots, L-1 \text{ and}$$

$$P[Y = L] = \sum_{k=L}^{\infty} \frac{H^k}{k!}e^{-H} = 1 - \Gamma_L(H)$$

where $\Gamma_L(H) = \frac{1}{\Gamma(L)}\int_H^{\infty} t^{L-1}e^{-t}dt$ is the upper incomplete Gamma function which is often used to derive theoretical results for QIS [123].

The statistical estimation of H based on Y can be carried out using the maximum-likelihood estimation. In the case of 1-bit measurements with $L = 1$, the random variable Y follows a Bernoulli distribution. The maximum-likelihood estimate is therefore found by maximizing the likelihood function of a sequence of independent Bernoulli random variables

$$\hat{H} = \arg\max \prod_{n=1}^{N} (1 - e^{-H})^{Y_n}(e^{-H})^{1-Y_n} = -\log(1 - \bar{Y})$$

where $\bar{Y}$ is the average of the sequence $\{Y_1, \ldots, Y_N\}$. For multibit signals, the maximum-likelihood estimation does not have a closed form. The typical workaround here is to first evaluate the statistical expectation $E[Y]$ (which is a function $\mu(.)$ of the exposure H)

$$\mu(H) = E[Y] = H\Gamma_{L-1}(H) + L(1 - \Gamma_L(H))$$

and construct the estimate as the functional inverse of μ

$$\hat{H} = \mu^{-1}(E[Y]).$$

Estimators constructed in such a way satisfy the so-called mean invariance property [125].