Modeling the photon counting and photoelectron counting characteristics of quanta image sensors

Bowen Liu and Jiangtao Xu†

Tianjin Key Laboratory of Imaging and Sensing Microelectronic Technology, School of Microelectronics, Tianjin University, Tianjin 300072, China

Abstract: A signal chain model of single-bit and multi-bit quanta image sensors (QISs) is established. Based on the proposed model, the photoresponse characteristics and signal error rates of QISs are investigated, and the effects of bit depth, quantum efficiency, dark current, and read noise on them are analyzed. When the signal error rates towards photons and photoelectrons counting are lower than 0.01, the high accuracy photon and photoelectron counting exposure ranges are determined. Furthermore, an optimization method of integration time to ensure that the QIS works in these high accuracy exposure ranges is presented. The trade-offs between pixel area, the mean value of incident photons, and integration time under different illuminance level are analyzed. For the 3-bit QIS with 0.16 e-/s dark current and 0.21 e- r.m.s. read noise, when the illuminance level and pixel area are 1 lux and 1.21 μm^2, or 10 000 lux and 0.21 μm^2, the recommended integration time is 8.8 to 30 ms, or 10 to 21.3 μs, respectively. The proposed method can guide the design and operation of single-bit and multi-bit QISs.

Key words: CMOS image sensor; quanta image sensor; photon counting; photoelectron counting; signal error rate; integration time

1. Introduction

The quanta image sensor (QIS)[1] is a modified CMOS image sensor (CIS)[2] that contains a large number of photon-counting pixels, and oversamples both spatially and temporally. The specialized pixels in QISs, which have sub-diffraction-limit (SDL) size, are referred to as "jots." Based on the bit depth of analog-to-digital converters (ADCs), QISs are categorized as single-bit QISs and multi-bit QISs.

To intuitively comprehend the concept of the QIS, a conceptual illustration is shown in Fig. 1. As shown in Fig. 1, the output of the QIS is a large spatial-temporal data cube that can be flexibly combined in subsequent image processing.

The main development direction of image sensors is high integration, high performance, and low cost. Therefore, the size of active pixels in CISs keeps shrinking to achieve higher resolution on a smaller chip size, which meets the needs of high integration and low cost. Nevertheless, the shrinkage of pixels causes the problem of full well capacity (FWC) decrease that leads to signal-to-noise ratio (SNR) and dynamic range (DR) decreases[3], which limits the further development of CISs to high performance.

To alleviate the problem of FWC decrease, the QIS, which is a candidate for the next-generation solid image sensor, was first introduced in 2005[4] and was then presented in 2011[5]. Progress has been made in theoretical research and engineering realization of QISs in recent years. The required conditions to achieve photon and photoelectron counting based on active pixel image sensors were analyzed theoretically in

Refs. [6, 7]. A theoretical model of signal, noise, photoresponse, SNR, and DR for single-bit and multi-bit QISs was established in Refs. [8, 9]. The bit error rate (BER) as a function of the exposure level and read noise for QISs was analyzed in Ref. [10], and the recommended input-referred read noise was lower than 0.15 e- r.m.s.. A modified 1/*f* noise model of the source follower in jots, which was fitted to the measured data, was established in Ref. [11]. A pump-gate jot with high conversion gain (CG) of 423 μV/e- and low read noise of 0.22 e- r.m.s. was presented in Refs. [12, 13]. A low-power readout circuit with a modified charge-transfer amplifier (CTA) was proposed in Ref. [14]. Moreover, a 1 Mjot 1040 fps QIS chip in the stacked backside-illuminated (BSI) process was reported in Refs. [15, 16].

The prospective application fields of QISs include consumer imaging, low-light imaging, security monitoring imaging, scientific imaging, medical imaging, and space imaging. When the image performance strongly depends on the absolute measurements of incident photons, especially in scientific and medical imaging, high accuracy photon and photoelectron counting is required. Therefore, the high accuracy photon and photoelectron counting characteristics of QISs should be studied further.

The photoresponse characteristic of QISs, which is referred as *D*–log*H* response, was theoretically derived in Ref. [9] and experimentally proven in Ref. [17]. On the one hand, for high dynamic range (HDR) applications, the overexposure latitude range of the *D*–log*H* response curve can be used to extend the DR of the QIS. On the other hand, for applications that need high accuracy photon and photoelectron counting, the non-linear response of overexposure latitude range will increase the counting error of the QIS in high light level, and the non-ideal characteristics, such as dark cur-
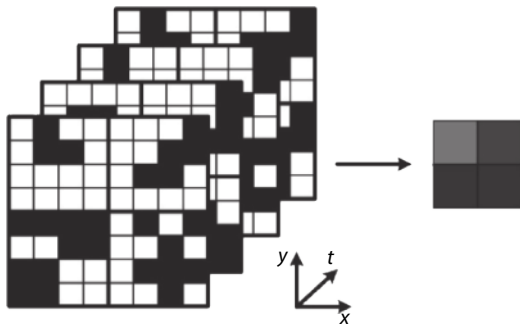
Fig. 1. The QIS conceptual illustration. An $8 \times 8 \times 4$ spatial-temporal data cube of jots in the QIS (left) is reconstructed to a 2×2 data plane of pixels in the output image (right). Each data of pixels is equal to the sum of a $4 \times 4 \times 4$ data sub-cube of jots.
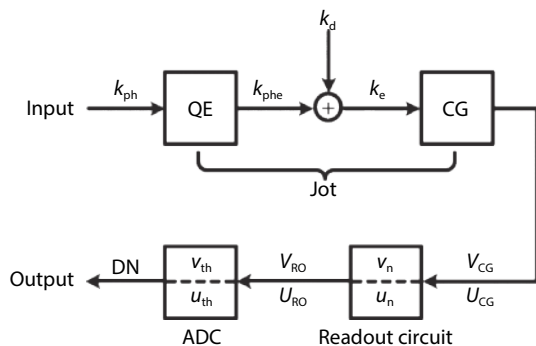

Fig. 2. The signal chain model of quanta image sensors.

Table 1.   The parameters of the signal chain model in Fig. 2.

Symbol	Parameter	Unit
k_{ph}	Number of photons	photon, p
k_{phe}	Number of photoelectrons	electron, e-
k_d	Number of dark signal electrons	e-
k_e	Number of total signal electrons	e-
V_{CG}	Voltage-referred jot output	microvolt, μV
U_{CG}	Electron-referred jot output	e-
V_{RO}	Voltage-referred readout circuit output	μV
U_{RO}	Electron-referred readout circuit output	e-
DN	ADC output digital numbers	ADU
QE	Quantum efficiency of the jot	e-/p
CG	Conversion gain of the jot	μV/e-
v_n	Voltage-referred read noise of the readout circuit	μV r.m.s.
u_n	Electron-referred read noise of the readout circuit	e- r.m.s.
v_{th}	Voltage-referred quantizer threshold of the ADC	μV
u_{th}	Electron-referred quantizer threshold of the ADC	e-

the jot, k_{ph} incident photons arrive at the pinned photodiode and generate k_{phe} photoelectrons on the condition that the quantum efficiency of the jot is QE. Then, k_{phe} photoelectrons and k_d dark signal electrons, which are generated from dark current, constitute together k_e total signal electrons. Then, k_e total signal electrons are transferred to the floating diffusion and converted to V_{CG} on the condition that the conversion gain of the jot is CG. In the readout circuit, V_{CG} is converted to V_{RO} on the condition that the read noise of the readout circuit is v_n. In the ADC, V_{RO} is converted to DN on the condition that the quantizer threshold of the ADC is v_{th}.

It is noted that the electron-referred values (i.e., the normalized values) of V_{CG}, V_{RO}, v_n, and v_{th} are U_{CG}, U_{RO}, u_n, and u_{th}, respectively. In most cases, the normalized values are introduced because it is more convenient to use them in mathematical derivation. The unit of ro (i.e., ADU) is equivalent to p or e-, when the ADC output digital numbers are compared with photons or photoelectrons. To analyze the proposed signal chain model mathematically, a series of probability distributions are introduced in the following sections.

rent and read noise, will increase the counting error of the QIS[10] in low light level.

Therefore, the photoresponse and counting error characteristics of QISs, and the effects of non-linear response and non-ideal characteristics to them are needed to be investigated. The high accuracy photon and photoelectron counting exposure ranges of QISs are required to be determined theoretically for subsequent on-chip or off-chip calibrations, and the integration time optimization under different light level and jot size is one of the common methods.

In this paper, an integration time optimization method towards high accuracy photon and photoelectron counting is presented to guide the design and operation of single-bit and multi-bit QISs. The remainder of this paper is organized as follows. A signal chain model of single-bit and multi-bit QISs is established in Section 2. The photoresponse characteristics for ideal and realistic QISs are investigated in Section 3. The signal error rates towards photon and photoelectron counting are investigated in Section 4. An optimization method of integration time that can guide the design and operation of QISs is presented in Section 5. Finally, the conclusions are given in Section 6.

2. Signal chain model

2.1. Basic concepts

The signal chain model of single-bit and multi-bit QISs is illustrated in Fig. 2, which contains signal conversion processes of photons to electrons, electrons to voltages, and voltages to digital numbers. The parameters of the signal chain model in Fig. 2 are listed in Table 1.

As shown in Fig. 2 and Table 1, the input and output signals of the proposed model are k_{ph} and DN, respectively. In

2.2. Total signal electrons

In the proposed signal chain model, the total signal electrons consist of photoelectrons and dark signal electrons, which are generated from incident photons and dark current in the jot, respectively. As described in Refs. [13, 18], benefiting from the jot device structure and the QIS operating characteristic, the image lag and crosstalk of pump-gate jots in the QIS are acceptable for photon counting and photoelectron counting applications under different exposure levels. Therefore, other sources of total signal electrons, including image lag, optical crosstalk, and electronic crosstalk, are not considered in the proposed model.

The emission of photons from most light sources is well-described by Poisson distribution[9]. Thus, over the integration time τ, the probability mass function (PMF) that k_{ph} incident photons arrive at the jot is defined as:

$$p[k_{ph}] = \frac{\mu_{ph}^{k_{ph}}}{k_{ph}!} e^{-\mu_{ph}}, \qquad (1)$$

where μ_{ph} is the mean value of incident photons, and k_{ph} is a nonnegative integer.

The quantum efficiency of the jot QE is defined as the number of photoelectrons divided by the number of incident photons. The process that photons convert to photoelectrons obeys binomial distribution[7]. Thus, the PMF that k_{phe} photoelectrons generate from k_{ph} incident photons under QE is defined as:

$$p[k_{phe}; k_{ph}, QE] = \frac{k_{ph}!}{k_{phe}!(k_{ph}-k_{phe})!}QE^{k_{phe}}(1-QE)^{k_{ph}-k_{phe}},$$

(2)

where k_{phe} and k_{ph} are nonnegative integers, and k_{phe} is less than or equal to k_{ph}.

Then, the PMF that k_{phe} photoelectrons are generated in the jot under μ_{ph} and QE, which is the convolution of two mutually independent PMFs Eqs. (1) and (2), is derived as:

$$p[k_{phe}] = \sum_{k_{ph}=k_{phe}}^{\infty} p[k_{ph}]p[k_{phe}; k_{ph}, QE]$$

(3)

$$= \frac{(QE \cdot \mu_{ph})^{k_{phe}}}{k_{phe}!}e^{-\mu_{ph}}\sum_{k_{ph}=k_{phe}}^{\infty}\frac{1}{(k_{ph}-k_{phe})!}[(1-QE)\mu_{ph}]^{k_{ph}-k_{phe}}$$

(4)

$$= \frac{(QE \cdot \mu_{ph})^{k_{phe}}}{k_{phe}!}e^{-\mu_{ph}}e^{(1-QE)\mu_{ph}}$$

(5)

$$= \frac{(QE \cdot \mu_{ph})^{k_{phe}}}{k_{phe}!}e^{-QE\cdot\mu_{ph}}$$

(6)

$$= \frac{\mu_{phe}^{k_{phe}}}{k_{phe}!}e^{-\mu_{phe}},$$

(7)

where $p[k_{phe}]$ obeys Poisson distribution, and μ_{phe} is the mean value of the photoelectrons.

Similarly, the emission of dark signal electrons that generate from dark current in active pixels also obeys Poisson distribution. Thus, over the integration time τ, the PMF that k_d dark signal electrons are generated in the jot is defined as:

$$p[k_d] = \frac{\mu_d^{k_d}}{k_d!}e^{-\mu_d},$$

(8)

where μ_d is the mean value of dark signal electrons, and k_d is a nonnegative integer.

Then, the PMF that k_e total signal electrons are generated in the jot, which is the convolution of two mutually independent PMFs Eqs. (7) and (8), is derived as:

$$p[k_e] = \sum_{k_{phe}=0}^{k_e} p[k_{phe}]p[k_d = k_e - k_{phe}]$$

(9)

$$= \frac{1}{k_e!}e^{-(\mu_{phe}+\mu_d)}\sum_{k_{phe}=0}^{k_e}\frac{k_e!}{k_{phe}!(k_e-k_{phe})!}\mu_{phe}^{k_{phe}}\mu_d^{k_e-k_{phe}}$$

(10)

$$= \frac{(\mu_{phe}+\mu_d)^{k_e}}{k_e!}e^{-(\mu_{phe}+\mu_d)}$$

(11)

$$= \frac{\mu_e^{k_e}}{k_e!}e^{-\mu_e},$$

(12)

where $p[k_e]$ also obeys Poisson distribution, μ_e is the mean value of total signal electrons, and k_e is a nonnegative integer.

2.3. Normalized jot output

The total signal electrons are transferred from pinned photodiode (PPD) to floating diffusion (FD) and converted to V_{CG} in the jot, when the conversion gain of the jot is CG that is calculated as:

$$CG = \frac{q}{C_{FD}},$$

(13)

where q is the electron charge, C_{FD} is the total capacitance of FD, and the unit of CG is usually $\mu V/e$-.

Then, the voltage-referred jot output is defined as:

$$V_{CG} = CG \cdot k_e.$$

(14)

It is convenient to use normalized values to simplify the analysis of the proposed model. Thus, the normalized value (i.e., the electron-referred value) of V_{CG} is calculated as[10]:

$$U_{CG} = \frac{V_{CG}}{CG}.$$

(15)

In the subsequent derivation, a similar method is used to convert the voltage-referred values V_{RO}, v_n, and v_{th} to the normalized values U_{RO}, u_n, and u_{th}, respectively.

2.4. Normalized readout circuit output

The readout circuit in the proposed model contains the source follower (SF), the amplifying circuit, and the correlation double sampling (CDS) circuit before the ADC. The voltage-referred read noise of the readout circuit v_n, which contains thermal noise and $1/f$ noise, is normalized as[10]:

$$u_n = \frac{v_n}{CG}.$$

(16)

It is noted that only temporal noise is considered in the proposed model. In other words, the spatial noise (i.e., fixed pattern noise (FPN), including dark signal nonuniformity (DSNU) and photoresponse nonuniformity (PRNU)), which may cause by QE, CG, u_n, and u_{th} nonuniformities among jots array, are not considered in the proposed model.

The main approach of the QIS to achieve photon counting and photoelectron counting is to increase the conversion gain of jots to reduce the electron-referred read noise lower than 0.5 e- r.m.s. (i.e., deep sup-electron read noise level). For example, when v_n is equal to 100 μV r.m.s., CG is required to be higher than 200 $\mu V/e$-, i.e., C_{FD} is required to be lower than 0.8 fF, to make sure u_n is lower than 0.5 e- r.m.s..

When the read noise u_n obeys Gaussian distribution, the component probability density function (PDF) that the normalized readout circuit output U_{RO} generates from the normalized jot output U_{CG} is defined as[10]:

$$P[U_{RO}; U_{CG}] = \frac{1}{\sqrt{2\pi u_n^2}}\exp\left[-\frac{(U_{RO}-U_{CG})^2}{2u_n^2}\right].$$
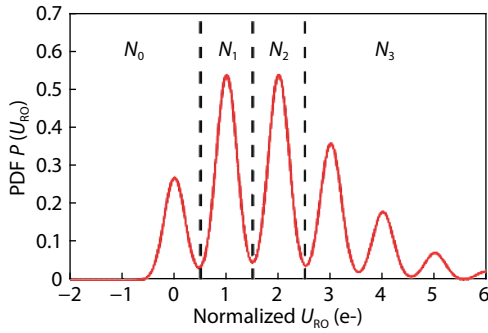
(17)

Fig. 3. (Color online) The PDF $P[U_{RO}]$ as a function of U_{RO} on the condition that μ_e = 2 e- and u_n = 0.2 e- r.m.s.. N_0, N_1, N_2, and N_3 are four quantization levels corresponding to 0, 1, 2, and 3 ADU for a 2-bit QIS. The dashed lines are three quantization boundaries corresponding to 0.5, 1.5, and 2.5 e- for a 2-bit QIS.

Using Eqs. (12), (14), (15), and (17), the PDF that the normalized readout circuit output U_{RO} is generated, when the mean value of total signal electrons is μ_e, is derived as:

$$P[U_{RO}] = \sum_{k_e=0}^{\infty} \frac{p[k_e]}{\sqrt{2\pi u_n^2}} \exp\left[-\frac{(U_{RO} - k_e)^2}{2u_n^2}\right], \quad (18)$$

where U_{RO} is a real number.

The PDF $P[U_{RO}]$ as a function of U_{RO} on the condition that μ_e = 2 e- and u_n = 0.2 e- r.m.s. is shown in Fig. 3 as an example.

2.5. ADC output

In the process of quantization, the normalized readout circuit output U_{RO} is converted to the ADC output DN based on different quantization levels. For an n-bit QIS, there are 2^n quantization levels and $2^n - 1$ quantization boundaries. The effective full well capacity of the QIS is generally limited by the bit depth of ADC and equal to $2^n - 1$.

As shown in Fig. 3, for a 2-bit QIS, four quantization levels are N_0, N_1, N_2, and N_3, which correspond to 0, 1, 2, and 3 ADU, respectively. When the normalized quantizer threshold u_{th} = 0.5 e-, three normalized quantization boundaries are 0.5, 1.5, and 2.5 e-. When $U_{RO} \in (-\infty, 0.5$ e-], DN = 0 ADU; when $U_{RO} \in (0.5$ e-, 1.5 e-], DN = 1 ADU; when $U_{RO} \in (1.5$ e-, 2.5 e-], DN = 2 ADU; when $U_{RO} \in (2.5$ e-, +∞], DN = 3 ADU.

Thus, for an n-bit QIS, the PMF that the ADC output digital numbers DN is generated, when the normalized quantizer threshold is u_{th}, is defined as:

$$p[DN] = \begin{cases} \int_{-\infty}^{U_{th1}} P[U_{RO}] dU_{RO}, & DN = 0, \\ \int_{U_{th2}}^{U_{th1}} P[U_{RO}] dU_{RO}, & 0 < DN < FWC_{eff}, \\ \int_{U_{th2}}^{+\infty} P[U_{RO}] dU_{RO}, & DN = FWC_{eff}, \\ 0, & DN > FWC_{eff}, \end{cases} \quad (19)$$

where the higher quantization boundary U_{th1} = DN + u_{th}, the lower quantization boundary U_{th2} = DN – (1 – u_{th}), the effective full well capacity FWC_{eff} = $2^n - 1$, and DN is a nonnegative integer.

3. Photoresponse characteristic

Conventional CISs generally have linear response, and some CISs can achieve logarithmic and linear-logarithmic re-

sponse through pixel or circuit structure design. In contrast from these CISs, the QIS has a special photoresponse characteristic, which is referred to as D–log$H$ response. For high dynamic range applications, the overexposure latitude range of the D–log$H$ response curve is used to extend the dynamic range of the QIS[9]. For applications that need high accuracy photon counting and photoelectron counting, the high accuracy linear range of the D–log$H$ response curve is required to be determined in theory.

3.1. Ideal D–log$H$ photoresponse curves

As shown in Fig. 1, by reconstructing a specified spatial-temporal data cube of jots into a single data point of a pixel in the output image, the trade-off between the equivalent full well capacity, spatial resolution, and temporal resolution can be flexibly achieved. The spatial and temporal oversampling coefficients of the QIS are defined as M_S and M_T, respectively, and the total oversampling coefficient of the QIS is calculated as:

$$M = M_S M_T, \quad (20)$$

where M has various spatial-temporal combinations. For example, when M = 512, a 4 × 4 × 32, 8 × 8 × 8, 16 × 16 × 2, et al. sub-cube of jots is alternative.

Assuming the jots used for reconstruction are completely equivalent (i.e., no DSNU and PRNU among jots array), the signal value of ADC output is calculated as:

$$S_{DN} = \sum_{i=0}^{M} S_{DN,i} = M \cdot \sum_{DN=0}^{\infty} DN \cdot p[DN], \quad (21)$$

where $p[DN]$ is given by Eq. (19).

Similarly, the saturated signal value of ADC output is calculated as:

$$S_{DN,sat} = FWC_{equ} = M \cdot FWC_{eff} = M(2^n - 1), \quad (22)$$

where FWC_{equ} and FWC_{eff} is the equivalent and effective full well capacity respectively, and n is the bit depth of QISs (i.e., the bit depth of ADCs in those QISs). For example, for a 3-bit QIS, when M_S = 4 × 4, M_T = 8 and M = 4 × 4 × 8 = 128, FWC_{equ} = $M \cdot FWC_{eff}$ = 896 e-.

In Ref. [9], in the D–log$H$ function, H denotes quanta exposure, which is defined as the average number of photons or photoelectrons arrive at a jot over the integration period; and D denotes bit density, which is defined as the number of saturated jots divided by the number of total jots for 1-bit QISs.

In this paper, the D–log$H$ function is extended to n-bit QISs. H denotes quanta exposure, which is defined as the mean value of photons that arrive at a jot over the integration time τ. Thus, H is equal to μ_{ph}. When QE = 1 e-/p, H is also equal to $\mu_{phe} \times 1$ p/e-. Meanwhile, D denotes signal density, which is defined as:

$$D = \frac{S_{DN}}{S_{DN,sat}} = \frac{\sum_{DN=0}^{\infty} DN \cdot p[DN]}{2^n - 1}, \quad (23)$$

where S_{DN} and $S_{DN,sat}$ are given by Eqs. (21) and (22) respectively.

The ideal D-log$H$ response curves for 1-bit, 2-bit, 3-bit, 4-bit, and 5-bit QISs are shown in solid lines in Fig. 4, where
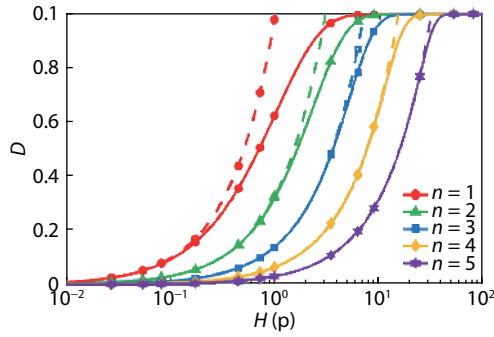
Fig. 4. (Color online) The ideal D-log$H$ response curves for 1-bit to 5-bit QISs in solid lines. And the ideal linear response curves for 1-bit to 5-bit CISs in dashed lines.

Table 2. Different conditions for realistic QISs in Figs. 5, 7, and 8.

Number	QE (e-/p)	μ_d (e-)	u_n (e- r.m.s.)
Curve 1	1	0	0
Curve 2	0.8	0	0
Curve 3	1	0.01	0
Curve 4	1	0	0.3
Curve 5	0.8	0.01	0.3

QE = 1 e-/p, μ_d = 0 e-, u_n = 0 e- r.m.s., and u_{th} = 0.5 e-. The ideal linear response curves for 1-bit to 5-bit CISs are shown in dashed lines in Fig. 4, where the output signal is equal to the input signal (i.e., the ideal accuracy photon counting is achieved).

As shown in Fig. 4, when D = 0.99, the saturated quanta exposure H_{sat} of 1-bit to 5-bit QISs are equal to 4.7, 7.3, 13, 22, and 39 p, respectively. Meanwhile, H_{sat} of 1-bit to 5-bit CISs are equal to 0.99, 2.97, 6.93, 14.85, and 30.69 p, respectively. Thus, compared with the n-bit CIS, the n-bit QIS has a higher H_{sat} and a greater potential in HDR applications in theory.

As H increases from 0.01 to 100 p, the QIS gradually enters in the nonlinear response range (the non-overlap range of solid line and dashed line) from the linear response range (the overlap range of solid line and dashed line). As n increases from 1 to 5, the linear and nonlinear response ranges of 1-bit to 5-bit QISs are extended and narrowed respectively. Thus, QISs with higher bit depth are inferred to have a better performance in applications that need high accuracy photon counting and photoelectron counting in theory.

3.2. Realistic photoresponse curves

For realistic QISs, QE < 1 e-/p, μ_d > 0 e-, u_n > 0 e- r.m.s., and u_{th} = 0.5 e-. In the realistic response curve for an n-bit QIS, H is equal to μ_{ph}, and D is calculated as Eq. (23). The realistic response curves for the 3-bit QIS under five different conditions, which are listed in Table 2, are shown in Fig. 5.

Based on Fig. 5, the effects of QE, μ_d, and u_n to realistic response curves for the QIS are qualitatively analyzed, as follows:

(1) Curve 1 (ideal) is the ideal response curve for the 3-bit QIS as a comparison;

(2) Curve 2 (QE = 0.8 e-/p) can be approximatively considered to be obtained by shifting curve 1 to the right. This means that when QE < 1 e-/p, a higher H is required to get the same D in curve 1;
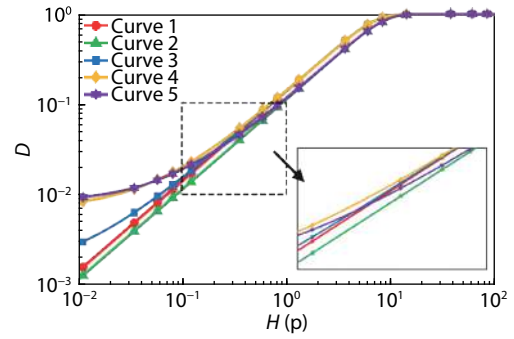

Fig. 5. (Color online) The realistic response curves for the 3-bit QIS. The different conditions of curve 1 to 5 are listed in Table 2.

(3) In curve 3 (μ_d = 0.01 e-), as H increases from 0.01 p to 100 p, curve 3 gradually coincides with curve 1, the difference of D between curve 3 and curve 1 gradually decreases, i.e., μ_d > 0 e- has an obvious effect on low light level (H < 0.1 p);

(4) In curve 4 (u_n = 0.3 e- r.m.s.), u_n > 0 e- r.m.s. has a similar effect on low light level (H < 0.1 p) compared with μ_d > 0 e- in curve 3;

(5) Curve 5 (QE = 0.8 e-/p, μ_d = 0.01 e-, and u_n = 0.3 e- r.m.s.) combines the effects of QE < 1 e-/p, μ_d > 0 e-, and u_n > 0 e- r.m.s.. When H is lower than 0.25 p, D in curve 5 is higher than curve 1, μ_d > 0 e- and u_n > 0 e- r.m.s. are dominant. When H is higher than 0.25 p, D in curve 5 is lower than curve 1, QE < 1 e-/p is dominant.

In summary, the lower boundary of the linear response range of the QIS is limited by μ_d and u_n, the higher boundary of the above range is limited by the bit depth of QISs n, and this range shifts to the right when QE < 1 e-/p. To further study the high accuracy linear ranges towards photon counting and photoelectron counting of ideal and realistic response curves, and the effects of QE, μ_d, and u_n to the above ranges of realistic response curves, a quantitative analysis is performed in Section 4.

4. Signal error rate

In this paper, the signal error rate is defined as the absolute value of the difference between the input and output signals divided by the input signal, which is used as the qualifying factor to evaluate the accuracy of photon counting or photoelectron counting for the QIS.

4.1. Signal error rate for ideal QISs

Assuming that the jots used for reconstruction are completely equivalent, like Eq. (21), the signal value of incident photons is calculated as:

$$S_{ph} = M \sum_{k_{ph}=0}^{\infty} k_{ph} \cdot p[k_{ph}] = M\mu_{ph}. \quad (24)$$

Similarly, the signal value of photoelectrons is calculated as:

$$S_{phe} = M \sum_{k_{phe}=0}^{\infty} k_{phe} \cdot p[k_{phe}] = M\mu_{phe}. \quad (25)$$

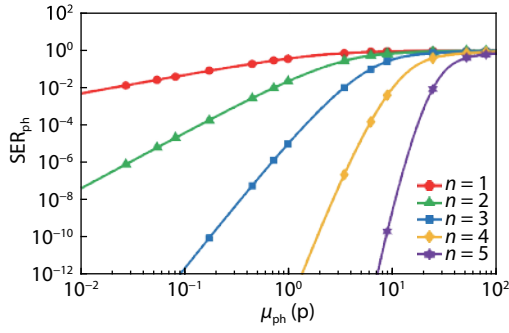And the signal error rate between S_{DN} and S_{ph} is defined as:

Fig. 6. (Color online) The photon counting signal error rate SER_{ph} as a function of the mean value of incident photons μ_{ph} for ideal 1-bit to 5-bit QISs.

$$\text{SER}_{\text{ph}} = \frac{|S_{\text{DN}} - S_{\text{ph}}|}{S_{\text{ph}}} = \frac{\left| \sum_{\text{DN}=0}^{\infty} \text{DN} \cdot p\,[\text{DN}] - \mu_{\text{ph}} \right|}{\mu_{\text{ph}}}, \qquad (26)$$

where SER_{ph} is the qualifying factor of evaluating the accuracy of photon counting in the QIS. When SER_{ph} is decreased, the difference between S_{DN} and S_{ph} is decreased, and the accuracy of photon counting is increased.

The signal error rate between S_{DN} and S_{phe} is defined as:

$$\text{SER}_{\text{phe}} = \frac{|S_{\text{DN}} - S_{\text{phe}}|}{S_{\text{phe}}} = \frac{\left| \sum_{\text{DN}=0}^{\infty} \text{DN} \cdot p\,[\text{DN}] - \mu_{\text{phe}} \right|}{\mu_{\text{phe}}}, \qquad (27)$$

where SER_{phe} is the qualifying factor of evaluating the accuracy of photoelectron counting in the QIS. When SER_{phe} is decreased, the difference between S_{DN} and S_{phe} is decreased, and the accuracy of photoelectron counting is increased.

In this paper, based on Eqs. (26) and (27), the μ_{ph} value interval corresponding to $\text{SER}_{\text{ph}} < 0.01$ is defined as the high accuracy photon counting exposure range (hereinafter to be referred as R_{ph}) or high accuracy linear range of response curves, and the μ_{phe} value interval corresponding to $\text{SER}_{\text{phe}} < 0.01$ is defined as the high accuracy photoelectron counting exposure range (hereinafter to be referred as R_{phe}).

As for n-bit ideal QISs, QE = 1 e-/p, μ_{d} = 0 e-, u_{n} = 0 e-r.m.s., and u_{th} = 0.5 e-. In this case, $\mu_{\text{phe}} = \mu_{\text{ph}} \times 1$ e-/p, and $\text{SER}_{\text{phe}} = \text{SER}_{\text{ph}}$, i.e., photon counting is equivalent to photoelectron counting when QE = 1 e-/p. And SER_{ph} as a function of μ_{ph} for ideal 1-bit to 5-bit QISs is shown in Fig. 6.

As shown in Fig. 6, for a given n, when μ_{ph} is increased, SER_{ph} is increased; for a given μ_{ph}, when n is increased, SER_{ph} is decreased; for ideal 1-bit to 5-bit QISs, R_{ph} is (−∞, 0.021 p), (−∞, 0.72 p), (−∞, 3.4 p), (−∞, 10 p), and (−∞, 25 p), respectively; in other words, as n increases from 1 to 5, R_{ph} is extended from (−∞, 0.021 p) to (−∞, 25 p).

Based on these results, it can be proven theoretically that increasing the bit depth of the QIS is an effective approach to improve the performance of QISs in applications that need high accuracy photon counting and photoelectron counting.

4.2. Photon counting signal error rate for realistic QISs

As for n-bit realistic QISs, QE < 1 e-/p, μ_{d} > 0 e-, u_{n} > 0 e-r.m.s., and u_{th} = 0.5 e-. In this case, photon counting is inequivalent to photoelectron counting. To determine the R_{ph} of realistic QISs, SER_{ph} as a function of μ_{ph} for the 3-bit QIS under

Fig. 7. (Color online) The photon counting signal error rate SER_{ph} as a function of the mean value of incident photons μ_{ph} for the 3-bit QIS. The five conditions of curve 1 to 5 are listed in Table 2.

five different conditions, which are listed in Table 2, is shown in Fig. 7.

Based on Fig. 7, the R_{ph} of curve 1 to 5 is (−∞, 3.4 p), N/A, (0.99 p, 3.6 p), (1.2 p, 3.4 p), and (0.23p, 0.26p), respectively. And the effects of QE, μ_{d}, and u_{n} to the R_{ph} of the realistic QIS are quantitative analyzed as follows:

(1) Curve 1 (ideal) is the ideal SER_{ph} curve for the 3-bit QIS as a comparison;

(2) In curve 2 (QE = 0.8 e-/p), when $\mu_{\text{ph}} \leqslant 1.6$ p, SER_{ph} remains constant at 0.2; when $\mu_{\text{ph}} > 1.6$ p, SER_{ph} is gradually increased to 1. This means that QE has significant effect on SER_{ph}, and QE is the main limit factor of high accuracy photon counting for the QIS;

(3) In curve 3 (μ_{d} = 0.01 e-), as μ_{ph} increases from 0.01 p to 100 p, SER_{ph} first decreases from a high value to the minimum value and then gradually increases to 1. The minimum value point (i.e., the optimal point for photon counting) is (μ_{ph} = 2.7 p, SER_{ph} = 2.23 × 10⁻⁴). Compared with curve 1, μ_{d} > 0 e- mainly increases the SER_{ph} on low light level (H < 0.1 p);

(4) In curve 4 (u_{n} = 0.3 e- r.m.s.), u_{n} > 0 e- r.m.s. has a similar effect on SER_{ph} compared with μ_{d} > 0 e- in curve 3. The minimum value point of curve 4 is (μ_{ph} = 2.4 p, SER_{ph} = 2.14 × 10⁻⁴);

(5) In curve 5 (QE = 0.8, μ_{d} = 0.01 e-, and u_{n} = 0.3 e-r.m.s.), the effects of QE < 1 e-/p, μ_{d} > 0 e-, and u_{n} > 0 e-r.m.s. are combined. As μ_{ph} increases from 0.01 p to 100 p, SER_{ph} first decreases and then increases. The minimum value point of curve 5 is (μ_{ph} = 0.24 p, SER_{ph} = 4.37 × 10⁻³). When μ_{ph} < 0.13 p, SER_{ph} > 0.2, the effects of μ_{d} = 0.01 e- and u_{n} = 0.3 e- r.m.s. are dominant; when 0.13 p $\leqslant \mu_{\text{ph}} \leqslant$ 3.5 p, 4.37 × 10⁻³ $\leqslant \text{SER}_{\text{ph}} \leqslant$ 0.2, the effect of QE = 0.8 e-/p is dominant; when μ_{ph} > 3.5 p, SER_{ph} is gradually increased to 1 due to the saturation of jots.

The cause of optimal point in curve 5 can be explained through the following example: assuming that 10 photons arrive at the jot and convert to 9 photoelectrons under QE; if 1 dark signal electron is generated, then the number of total signal electrons is 10; thus, a dark signal electron may fill the electron loss due to QE and cause SER_{ph} decrease in some special cases.

In summary, when QE is low, R_{ph} is limited by QE; when QE is high enough, such as 0.95 e-/p[7], the lower boundary of R_{ph} is limited by μ_{d} and u_{n}, and the higher boundary of R_{ph} is limited by the saturated quanta exposure (i.e., the bit depth of QISs). Therefore, the main approach of extending the R_{ph} of realistic QISs is increasing QE and n, and decreas-
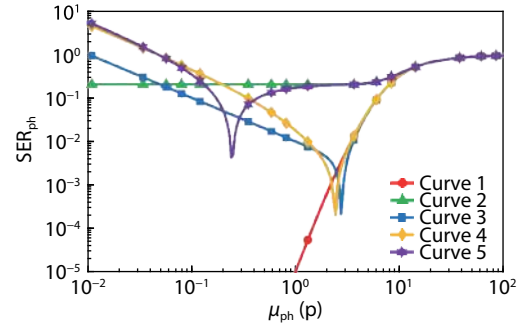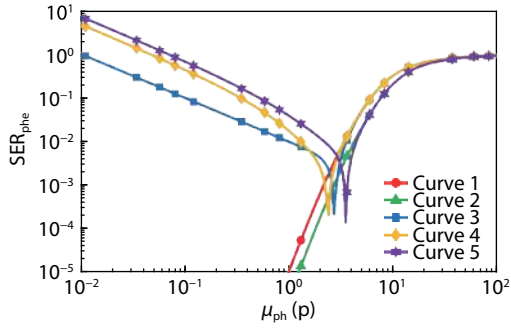
Fig. 8. (Color online) The photoelectron counting signal error rate SER_{phe} as a function of the mean value of incident photons μ_{ph} for the 3-bit QIS. The five conditions of curve 1 to 5 are listed in Table 2.

ing μ_d and u_n.

4.3. Photoelectron counting signal error rate for realistic QISs

To determine the R_{phe} of realistic QISs, SER_{phe} as a function of μ_{ph} for the 3-bit QIS under five different conditions, which are listed in Table 2, is shown in Fig. 8.

As shown in Fig. 8, the R_{phe} of curve 1 to 5 is $(-\infty, 3.4 \text{ p})$, $(-\infty, 4.3 \text{ p})$, $(0.99 \text{ p}, 3.6 \text{ p})$, $(1.2 \text{ p}, 3.4 \text{ p})$, and $(2.1 \text{ p}, 4.5 \text{ p})$, respectively. It is noted that since QE = 1 e-/p and SER_{phe} = SER_{ph}, curve 1, 3, and 4 in Fig. 8 are the same as Fig. 7.

Curve 2 (QE = 0.8 e-/p) can be approximatively considered to be obtained by shifting curve 1 (ideal) to the right (i.e., R_{phe} shifts to the right when QE < 1 e-/p).

In curve 5 (QE = 0.8 e-/p, μ_d = 0.01 e-, and u_n = 0.3 e-r.m.s.), the effects of QE < 1 e-/p, μ_d > 0 e-, and u_n > 0 e- r.m.s. are combined. As μ_{ph} increases from 0.01 to 100 p, SER_{phe} first decreases and then increases. The minimum value point of curve 5 is (μ_{ph} = 3.5 p, SER_{phe} = 1.45 × 10^{-4}). When μ_{ph} < 3.5 p, SER_{phe} > 1.45 × 10^{-4}, the effects of μ_d = 0.01 e- and u_n = 0.3 e- r.m.s. are dominant; when μ_{ph} ⩾ 3.5 p, curve 5 gradually overlaps with curve 2, and SER_{phe} is finally increased to 1 due to the saturation of jots.

In summary, the lower boundary of R_{phe} is limited by μ_d and u_n, the higher boundary of R_{phe} is limited by the bit depth of QISs, and R_{phe} shifts to the right when QE < 1 e-/p. Therefore, the main approach of extending the R_{phe} of realistic QISs is increasing n, and decreasing μ_d and u_n.

5. Integration time optimization

5.1. Physical explanation

As discussed in Section 4, by calculating SER_{ph} and SER_{phe}, the μ_{ph} and μ_{phe} value interval corresponding to SER_{ph} < 0.01 and SER_{phe} < 0.01 (i.e., R_{ph} and R_{phe}) are determined.

In imaging applications, under the specified illuminance level and pixel size, by adjusting the integration time of the QIS, μ_{ph} and μ_{phe} can be adjusted to R_{ph} and R_{phe}, respectively, and the high accuracy photon counting and photoelectron counting can be achieved.

To propose the integration time optimization method, the illuminance of the light source I_{lux} should be converted to the mean value of incident photons μ_{ph} at first.

The conversion formula between I_{lux} and the mean value of photons that generate from the light source μ_s is defined as[19]:
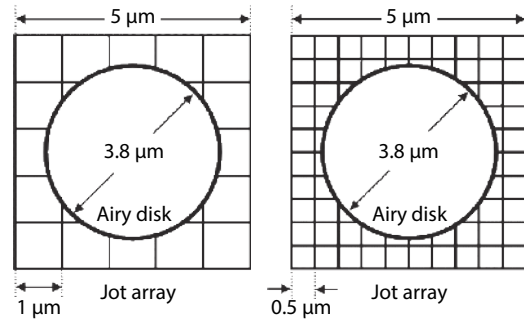

Fig. 9. The conceptual illustration of the Airy disk and jot array. The Airy disk diameter D_A = 3.8 μm, the jot area A_{jot} = 1 μm^2 (left) and A_{jot} = 0.25 μm^2 (right).

$$\mu_s = \frac{I_{lux}A_{jot}\tau}{KE_{ph}} = \frac{I_{lux}A_{jot}\tau\lambda}{Khc}, \tag{28}$$

where E_{ph} denotes the energy of a single photon, A_{jot} denotes the area of a single jot, τ denotes the integration time of jots, K denotes the conversion coefficient, h = 6.626 × 10^{-34} J·s denotes Planck constant, c = 2.998 × 10^8 m/s denotes the speed of light in vacuum, λ denotes the wavelength of photons.

The conversion formula between μ_s and μ_{ph} is defined as[19]:

$$\mu_{ph} = \frac{T(1-R) \cdot FF}{4F_N^2}\mu_s, \tag{29}$$

where F_N denotes the F number of lens, T denotes the transmittance of lens, R denotes the reflectance of the image sensor surface, and FF denotes the fill factor of the jots.

When λ = 555 nm, K = 683 lm·W^{-1} = 683 lux·m^2·W^{-1}[19], and let F_N = 2.8, T = 0.9, R = 0.1, and FF = 0.9. Using Eqs. (28) and (29), the conversion formula between I_{lux} and μ_{ph} is calculated as:

$$\mu_{ph} = LI_{lux}A_{jot}\tau, \tag{30}$$

where L is the corrected parameter of the specified QIS, and L = 9.51 × 10^{-5} lux^{-1}·μm^{-2}·μs^{-1} in this case.

For an ideal optical imaging system, the perfect lens focuses the point light source to a diffraction-limit spot, which is referred to as Airy disk, and the spot is surrounded by higher order diffraction rings. The diameter of Airy disk is calculated as[4]:

$$D_A = 2.44\lambda F_N. \tag{31}$$

When λ = 555 nm and F_N = 2.8, D_A = 3.8 μm. To intuitively describe the geometric dimensioning of the Airy disk and jot array, and make a physical explanation of the relationship between I_{lux}, A_{jot}, and τ, a conceptual illustration is shown in Fig. 9.

As shown in Fig. 9, when A_{jot} = 1 μm^2 or A_{jot} = 0.25 μm^2, the same Airy disk covers 9 or 44 jots, respectively. This means that under specified I_{lux} and τ, when A_{jot} is larger, the jot has a higher probability of collecting more incident photons. In other words, when I_{lux} and A_{jot} are equal to specified values, μ_{ph} or μ_{phe} can be adjusted to be within R_{ph} or R_{phe} by optimizing integration time τ.

Table 3.   The correlation parameters of the QIS chip in Ref. [16].

Parameter	Value
Jot size	$1.1 \times 1.1\ \mu m^2$
Quantum efficiency	79% at 550 nm
Dark current	0.16 e-/s/jot
Read noise	0.21 e- r.m.s.
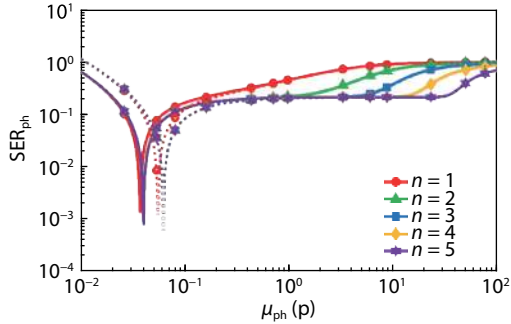
Fig. 10. (Color online) The photon counting signal error rate SER_{ph} as a function of the mean value of incident photons μ_{ph} for 1-bit to 5-bit QISs based on the parameters listed in Table 3.
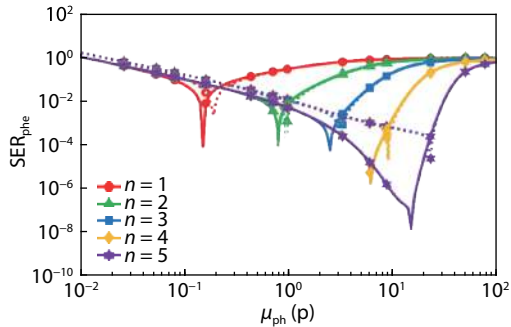
Fig. 11. (Color online) The photoelectron counting signal error rate SER_{phe} as a function of the mean value of incident photons μ_{ph} for 1-bit to 5-bit QISs based on the parameters listed in Table 3.

5.2. Optimization method

To further describe the integration time optimization method for QISs, the correlation parameters of the 1 Mjot QIS chip in Ref. [16] are listed in Table 3.

Based on the parameters listed in Table 3, QE = 0.79 e-/p, $\mu_d = 1.6 \times 10^{-6}$ e- to 4.8×10^{-3} e- (assuming $\tau = 10\ \mu s$ to 30 ms), $u_n = 0.21$ e- r.m.s., and $u_{th} = 0.5$ e-. SER_{ph} as a function of μ_{ph} is shown in Fig. 10, and SER_{phe} as a function of μ_{ph} is shown in Fig. 11. For solid lines, QE = 0.79 e-/p, $\mu_d = 1.6 \times 10^{-6}$ e-, $u_n = 0.21$ e- r.m.s., and $u_{th} = 0.5$ e-; for dashed lines, QE = 0.79 e-/p, $\mu_d = 4.8 \times 10^{-3}$ e-, $u_n = 0.21$ e- r.m.s., and $u_{th} = 0.5$ e-.

As shown in Fig. 10, for solid lines, R_{ph} is (0.035 p, 0.039 p) for 1-bit QIS, and R_{ph} is (0.038 p, 0.042 p) for 2-bit to 5-bit QISs. For dashed lines, R_{ph} is (0.052 p, 0.057 p) for 1-bit QIS, and R_{ph} is (0.059 p, 0.065 p) for 2-bit to 5-bit QISs.

As shown in Fig. 11, for solid lines, R_{phe} is (0.13 p, 0.17 p), (0.55 p, 1.1 p), (0.65 p, 4.3 p), (0.65 p, 13 p), and (0.65 p, 31 p) for 1 to 5 bit QISs, respectively. For dashed lines, R_{phe} is (0.17 p, 0.2 p), (0.74 p, 1.2 p), (1 p, 4.4 p), (1 p, 13 p), and (1 p, 31 p) for 1 to 5 bit QISs, respectively.

To sum up, as for the QISs in Figs. 10 and 11, QE is required to be further increased to extend R_{ph}; the dark cur-

rent in jots is small enough that μ_d has a slight effect on R_{phe} even for a long integration time τ; u_n is required to be further decreased to extend R_{phe}; and a higher bit depth of QISs n is suggested to extend R_{phe}.

As listed in Table 3, the jot area $A_{jot} = 1.21\ \mu m^2$. For a more advanced process, A_{jot} is predicted to be smaller in the future. The range of A_{jot} is assumed as 0.1 to 1.5 μm^2 in this paper. In addition, the acceptable integration time τ (hereinafter to be referred as R_τ) is assumed as 10 to 30 000 μs.

For the 3-bit QIS in Fig. 11, R_{phe} is (1 p, 4.4 p). Based on Eq. (30), the relationship between A_{jot}, μ_{ph}, and τ under different illuminance level I_{lux} is shown in Fig. 12.

As shown in Fig. 12, the area in orange denotes $\tau < 10\ \mu s$, and τ is below R_τ; the area in blue denotes $10\ \mu s \leqslant \tau \leqslant 30\ 000\ \mu s$, and τ is within the R_τ; the area in yellow denotes $\tau > 30\ 000\ \mu s$, and τ is above R_τ. As I_{lux} increases from 0.1 to 10 000 lux, the area in blue increases firstly and then decreases.

When I_{lux} is low, such as 0.1 lux, τ is required to reach a higher value that may be above R_τ; when I_{lux} is high, such as 10 000 lux, τ is required to reach a lower value that may be below R_τ. Thus, it will be more difficult to ensure that the QIS is working within R_{phe}, when the light level is too low or too high. Combined with the result in Section 4, when R_{ph} or R_{phe} is extended, the integration time optimization difficulty will be decreased. In addition, when I_{lux} is lower, a larger A_{jot} will decrease the optimization difficulty; when I_{lux} is higher, a smaller A_{jot} will decrease the optimization difficulty. For example, when $I_{lux} = 1$ lux and $A_{jot} = 1.21\ \mu m^2$, the recommended value interval of τ is (8.8 ms, 30 ms), and when $I_{lux} = 10\ 000$ lux and $A_{jot} = 0.21\ \mu m^2$, the recommended value interval of τ is (10 μs, 21.3 μs).

In summary, the trade-off between A_{jot}, μ_{ph}, and τ under different I_{lux} is needed to be considered, when the QIS is designed and operated. By integration time optimization, the QIS can work within R_{ph} or R_{phe}. The main approach of decreasing the optimization difficulty is to extend R_{ph} or R_{phe}, which required higher quantum efficiency, higher bit depth of QISs, lower dark current, and lower read noise. Therefore, when the correlation parameters of QISs are determined, the similar integration time optimization method can be performed to guide the design and operation of QISs.

6. Conclusion

In this paper, a signal chain model of single-bit and multi-bit QISs, which contains signal conversion processes of photons, electrons, voltages, and digital numbers, is established. Based on the proposed model, the photoresponse characteristics and signal error rates of ideal and realistic QISs are investigated. The effects of the ADC bit depth n, quantum efficiency QE, the mean value of dark signal electrons μ_d, and electron-referred read noise u_n to them are analyzed. The signal error rate between photons or photoelectrons and digital numbers SER_{ph} and SER_{phe} are defined as the quality factors of photon and photoelectron counting. When SER_{ph} and SER_{phe} are lower than 0.01, the high accuracy photon and photoelectron counting exposure ranges, which are referred as R_{ph} and R_{phe}, are determined. The main approach of extending R_{ph} and R_{phe} is increasing n and QE, and decreasing μ_d and u_n. Moreover, an integration time optimization method to en-
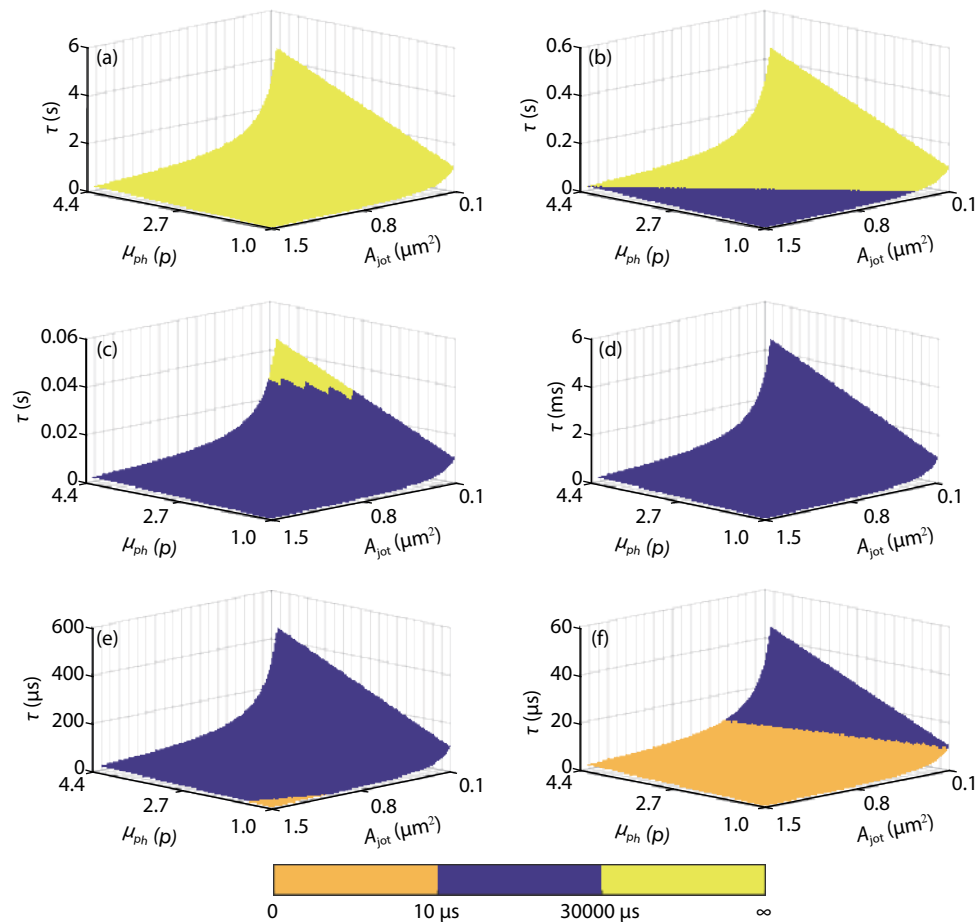
Fig. 12. (Color online) The relationship between jot area A_{jot}, the mean value of incident photons μ_{ph}, and integration time τ under different illuminance level I_{lux} for the 3-bit QIS in Fig. 11. (a) $I_{lux} = 0.1$ lux. (b) $I_{lux} = 1$ lux. (c) $I_{lux} = 10$ lux. (d) $I_{lux} = 100$ lux. (e) $I_{lux} = 1000$ lux. (f) $I_{lux} = 10\,000$ lux.

sure that the QIS works within R_{ph} and R_{phe} is presented. The trade-off between jot area A_{jot}, the mean value of incident photons μ_{ph}, and integration time τ under different illuminance level I_{lux} are analyzed. When n, A_{jot}, QE, μ_d, and u_n of QISs are determined, the integration time optimization method can be performed to guide the design and operation of QISs.

Acknowledgements

References

[1] Fossum E, Ma J J, Masoodian S, et al. The quanta image sensor: Every photon counts. Sensors, 2016, 16, 1260

[2] Fossum E R. CMOS image sensors: Electronic camera-on-a-chip. IEEE Trans Electron Devices, 1997, 44, 1689

[3] Fossum E R, Hondongwa D B. A review of the pinned photodiode for CCD and CMOS image sensors. IEEE J Electron Devices Soc, 2014, 2, 33

[4] Fossum E R. What to do with sub-diffraction-limit (SDL) pixels – A proposal for a gigapixel digital film sensor (DFS). Proceedings of Workshop on Charge-Coupled Devices and Advanced Image Sensors, 2005, 214

[5] Fossum E R. The quanta image sensor (QIS): Concepts and challenges. Proceedings of Imaging Systems and Applications, 2011, JTuE1

[6] Teranishi N. Toward photon counting image sensors. Proceedings of Imaging Systems and Applications, 2011, IMA1

[7] Teranishi N. Required conditions for photon-counting image sensors. IEEE Trans Electron Devices, 2012, 59, 2199

[8] Fossum E R. Application of photon statistics to the quanta image sensor. Proceedings of International Image Sensors Workshop, 2013, S9-1

[9] Fossum E R. Modeling the performance of single-bit and multi-bit quanta image sensors. IEEE J Electron Devices Soc, 2013, 1, 166

[10] Fossum E R. Photon counting error rates in single-bit and multibit quanta image sensors. IEEE J Electron Devices Soc, 2016, 4, 136

[11] Deng W, Starkey D, Ma J, et al. Modelling measured 1/f noise in quanta image sensor (QIS). Proceedings of International Image Sensors Workshop, 2019, R07

[12] Ma J J, Fossum E R. A pump-gate jot device with high conversion gain for a Quanta image sensor. IEEE J Electron Devices Soc, 2015, 3, 73

[13] Ma J J, Starkey D, Rao A R, et al. Characterization of quanta image sensor pump-gate jots with deep sub-electron read noise. IEEE J Electron Devices Soc, 2015, 3, 472

[14] Masoodian S, Rao A R, Ma J J, et al. A 2.5 pJ/b binary image sensor as a pathfinder for quanta image sensors. IEEE Trans Electron Devices, 2016, 63, 100

[15] Masoodian S, Ma J, Starkey D, et al. A 1Mjot 1040fps 0.22e- r. m. s. stacked BSI quanta image sensor with cluster-parallel readout. Proceedings of International Image Sensors Workshop, 2017, R19

[16] Starkey D, Deng W, Ma J, et al. Quanta imaging sensors: Achieving single-photon counting without avalanche gain. SPIE Defense + Security. Proceedings of Micro- and Nanotechnology Sensors, Systems, and Applications X, 2018, 1063, 106391Q

[17]   Dutton N, Parmesan L, Gnecchi S, et al. Oversampled ITOF imaging techniques using SPAD-based quanta image sensors. Proceedings of International Image Sensors Workshop, 2015, S6-4

[18]   Ma J J, Anzagira L, Fossum E R. A 1 μm-pitch quanta image sensor jot device with shared readout. IEEE J Electron Devices Soc, 2016, 4, 83

[19]   Jun O. Smart CMOS image sensors and applications. CRC Press, 2007, 181

Bowen Liu received the B.E. degree from the School of Physics and Optoelectronic Engineering, Dalian University of Technology, Dalian, China, in 2017. Since September 2017, he has been studying for the M.S. degree at School of Microelectronics, Tianjin University. His research interests include the high conversion gain pixel and low noise CMOS image sensor.

Jiangtao Xu received the B.E., M.S. and Ph.D. degrees from the School of Electronic Information and Engineering, Tianjin University, in 2001, 2004 and 2007, respectively. From 2007 to 2010 he was a Lecturer and from 2010 to 2018 he was an Associate Professor with the School of Electronic Information Engineering, Tianjin University. Since 2018, he has been a Professor at the School of Microelectronics, Tianjin University. His research interests include CMOS image sensor and image signal processing system.