Exposure-Referred Signal-to-Noise Ratio for Digital Image Sensors

Abhiram Gnanasambandam, *Student Member, IEEE*, and Stanley H. Chan, *Senior Member, IEEE*

Abstract—The signal-to-noise ratio (SNR) is a fundamental tool to measure the performance of an image sensor. However, confusions sometimes arise between the two types of SNRs. The first one is the output-referred SNR which measures the ratio between the signal and the noise seen at the sensor's output. This SNR is easy to compute, and it is linear in the log-log scale for most image sensors. The second SNR is the exposure-referred SNR, also known as the input-referred SNR. This SNR considers the noise at the input by including a derivative term to the output-referred SNR. The two SNRs have similar behaviors for sensors with a large full-well capacity. However, for sensors with a small full-well capacity, the exposure-referred SNR can capture some behaviors that the output-referred SNR cannot.

While the exposure-referred SNR has been known and used by the industry for a long time, a theoretically rigorous derivation from a signal processing perspective is lacking. In particular, while various equations can be found in different sources of the literature, there is currently no paper that attempts to assemble, derive, and organize these equations in one place. This paper aims to fill the gap by answering four questions: (1) How is the exposure-referred SNR derived from first principles? (2) Is the output-referred SNR a special case of the exposure-referred SNR, or are they completely different? (3) How to compute the SNR efficiently? (4) What utilities can the SNR bring to solving imaging tasks? New theoretical results are derived for image sensors of *any* bit-depth and full-well capacity.

Index Terms—Signal-to-noise ratio (SNR), full-well capacity, CMOS image sensors (CIS), charge coupled devices (CCD), quanta image sensors (QIS), single-photon imaging.

I. INTRODUCTION

The signal-to-noise ratio (SNR) is a basic tool to measure a device's performance when acquiring, transmitting, and processing raw data in the presence of noise. In as early as 1949, when Claude Shannon derived the information capacity of a noisy Gaussian channel, the concept of SNR was already presented [1]. As the name suggests, the SNR is the ratio between the signal power and the noise power

$$\text{SNR} = \frac{\text{signal power}}{\text{noise power}}, \tag{1}$$

which is sometimes expressed in the logarithmic scale via $10 \log_{10} \text{SNR}$ with the unit decibel (dB). Assuming that the signal follows the equation $Y = \theta + W$ where θ is a scalar and $W \sim \text{Gaussian}(0, \sigma^2)$ is the white noise, one can measure the SNR at the *output* by defining

$$\text{SNR}_{\text{out}}(\theta) = \frac{\text{signal at output}}{\text{noise at output}} = \frac{\mathbb{E}[Y]}{\sqrt{\text{Var}[Y]}}, \tag{2}$$

A. Gnanasambandam and S. Chan are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA. Email: {agnanasa, stanchan}@purdue.edu.

where the conversion from power to magnitude is taken care by changing the log from $10 \log_{10} \text{SNR}$ to $20 \log_{10} \text{SNR}_{\text{out}}$. In this equation, $\mathbb{E}[\cdot]$ denotes the expectation and $\text{Var}[\cdot]$ denotes the variance of the random variable Y. The SNR defined in (2) is known as the *output-referred* SNR.

In the sensor's literature, there is an alternative definition of the SNR which is based on the exposure (i.e., the input). The definition of this *exposure-referred* SNR is [2]

$$\text{SNR}_{\text{exp}}(\theta) = \frac{\text{signal at input}}{\text{noise at input}} = \frac{\theta}{\sqrt{\text{Var}[Y] \cdot \frac{d\theta}{d\mathbb{E}[Y]}}}. \tag{3}$$

The denominator of the equation is called the exposure-referred noise or the input-referred noise [3]. In computer vision, an early work of Mitsunaga and Nayar in 1999 [4] mentioned the equation (Equation (3)) although they did not give it a name.

The interpretation of the derivative $d\theta/d\mathbb{E}[Y]$ is known to the sensor's community [2]. The argument is that the input-referred noise is computed by back-propagating the signal from the output to the input via the transfer function $d\theta/d\mathbb{E}[Y]$ which approximates the relationship between the input θ and the output $\mathbb{E}[Y]$ [5, supplementary report]. As mentioned in multiple papers [5], [6], the presence of this derivative term is particularly important for sensors with a small full-well capacity because when a sensor saturates, the SNR has to drop. The exposure-referred SNR can capture this phenomenon whereas the output-referred SNR cannot.

Both versions of the SNRs are widely used in the industry [2]. People use the SNRs to evaluate sensors and to design sensor parameters so that the performance is maximized. However, from a theoretical point of view, besides the usual interpretation of treating the derivative as a transfer function, mathematically there is no rigorous proof trying to unify the two SNRs. The paper by Yang et al. [7] introduced the SNR from a statistical estimation perspective, whereas the paper by Elgendy and Chan [5] proved the equivalence between the SNR in [7] and SNR_{exp} for the case of one-bit sensors under a zero read noise. The connections between SNR_{exp} and SNR_{out} and their generalizations to multi-bit sensors remain unanswered. The goal of this paper is to fill the gap by answering four questions:

(i) How is the exposure-referred SNR derived from the first principle?

(ii) Under the unified framework addressed in (i), is it possible to show that SNR_{out} is a special case of SNR_{exp}?

(iii) How to efficiently compute SNR_{exp} for complex forward imaging models?

(iv) What are the utilities of $\text{SNR}_{\text{exp}}(\theta)$?

II. Mathematical Background

A. Limitations of Output-Referred SNR

To begin the discussion, it would be useful to first study the output-referred SNR and understand its limitations. Consider a Poisson-Gaussian random variable which is commonly used in modeling image sensors [8], [9]:

$$X \sim \text{Poisson}(\theta) + \text{Gaussian}(0, \sigma_{\text{read}}^2),$$

where θ denotes the average number of photons (i.e., the flux integrated over the surface area and the exposure time), and σ_{read} is the standard deviation of the read noise. This equation is a simplified model. Factors such as dark current, pixel non-uniformity, defects, analog-to-digital conversion, and color filter arrays are not considered to make the theoretical derivations tractable. Later in the paper, when the Monte Carlo simulation is introduced, some of these factors will be included.

Assuming that the sensor has a full-well capacity L (with the unit electrons), the output produced by the sensor follows the equation

$$Y = \begin{cases} X, & X < L, \\ L, & X \geq L. \end{cases} \quad (4)$$

If the full-well capacity is $L = \infty$, the measurement Y will never saturate, and hence the expectation is $\mathbb{E}[Y] = \theta$ and the variance is $\text{Var}[Y] = \theta + \sigma_{\text{read}}^2$. Then, according to (2), SNR_{out} can be computed via

$$\text{SNR}_{\text{out}}(\theta) = \frac{\theta}{\sqrt{\theta + \sigma_{\text{read}}^2}}. \quad (5)$$

If the read noise is negligible, i.e. $\sigma_{\text{read}} \ll \theta$, the SNR can be simplified to $\text{SNR}_{\text{out}}(\theta) = \sqrt{\theta}$. This equation can be found in many standard texts, e.g., [8].

The problem arises when the full-well capacity is *finite*. When $L < \infty$, SNR_{out} in (2) is still adequate to capture the sensor's behavior when the exposure θ is smaller than the full-well capacity L. However, if θ reaches the full-well and goes beyond it, the mean $\mathbb{E}[Y]$ will stop growing with θ as illustrated in Figure 1. The variance $\text{Var}[Y]$ will gradually drop to zero because when Y goes beyond the full-well capacity, it will be capped. As a result, SNR_{out} according to (2) will eventually go to infinity (because $\text{Var}[Y] \to 0$). This is not a desirable behavior because the SNR beyond the saturation is supposed to be poor.

B. One-bit and Multi-bit Sensors

If the full-well capacity is the source of the issue, the next question is how commonly does it happen. The full-well capacity of a conventional CMOS image sensor (CIS) is thousands of electrons. Therefore, unless the scene is bright and the integration is long, the sensor generally does not need to operate near the full-well capacity. However, for a newer type of image sensors based on single-photon detectors, the full-well capacity can be as small as only one electron. Their idea is to acquire binary bits at a very high frame rate and use computational methods to reconstruct the image. This class of sensors is broadly known as the quanta image sensors (QIS) that can be implemented using the single-photon avalanche
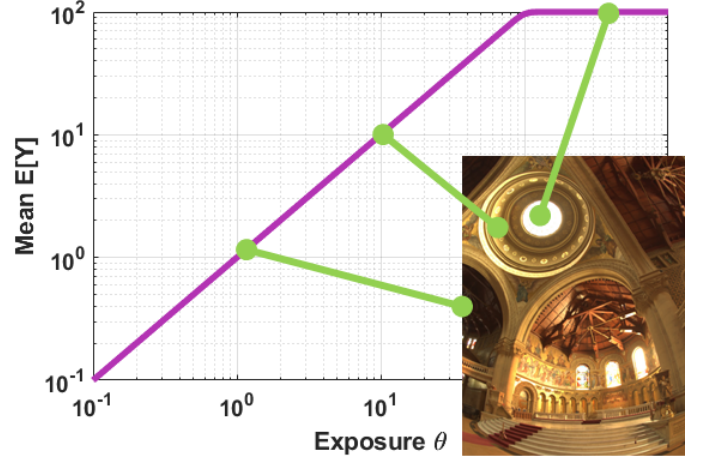

Fig. 1. With a finite full-well capacity, the mean $\mathbb{E}[Y]$ will stop growing when the exposure θ exceeds the full-well capacity $L = 10^2$.

diodes (SPAD-QIS) or the existing CMOS technology (CIS-QIS) [10].

From a mathematical modeling point of view, QIS is no different from a conventional CIS if it operates in the multi-bit mode because the equation will follow (4) [6]. The difference is that for QIS, the read noise is about 0.2 electrons whereas for CIS, the read noise can range from 2 to tens of electrons. If QIS operates in the one-bit mode, Y will follow

$$Y = \begin{cases} 0, & X < q, \\ 1, & X \geq q, \end{cases} \quad (6)$$

where q is the threshold (usually set as $q = 0.5$). Because of the generality of equations (4) and (6), the theoretical results of this paper is applicable to sensors of *any* bit-depth, including QIS and the conventional CIS.

C. Truncated Poisson and the Incomplete Gamma Function

The multi-bit sensor equation in (4) requires some statistical tools to handle the truncated Poisson random variable. To further simplify notations, in this section, the read noise is $\sigma_{\text{read}} = 0$ so that $X \sim \text{Poisson}(\theta)$. In this case, the probability mass function of Y is

$$p_Y(y) = \begin{cases} \frac{\theta^y}{y!} e^{-\theta}, & y < L, \\ \sum_{\ell=L}^{\infty} \frac{\theta^\ell}{\ell!} e^{-\theta}, & y = L. \end{cases} \quad (7)$$

By construction, the random variable Y will never take a value greater than L. The probability that $Y = L$ is given by the sum of the Poisson tail, which can be conveniently expressed via the incomplete Gamma function as shown in Figure 2.

Definition 1 (Incomplete Gamma function, [11]): The upper incomplete Gamma function is defined as $\Psi_L : \mathbb{R}_+ \to [0, 1]$, with

$$\Psi_L(\theta) = \frac{1}{\Gamma(L)} \int_\theta^\infty t^{L-1} e^{-t} dt = \sum_{\ell=0}^{L-1} \frac{\theta^\ell e^{-\theta}}{\ell!}, \quad (8)$$

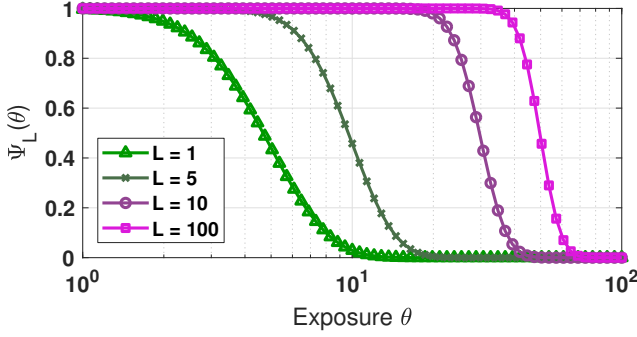for $\theta > 0, L \in \mathbb{N}$ where $\Gamma(L) = (L-1)!$ is the standard Gamma function.

Fig. 2. Incomplete Gamma function $\Psi_L(\theta)$ as a function of θ.

A few useful properties of $\Psi_L(\theta)$ can be derived. First, the first-order derivative of $\Psi_L(\theta)$ is

$$\Psi'_L(\theta) = -\frac{\theta^{L-1}e^{-\theta}}{(L-1)!} < 0, \quad \text{for all } \theta, \qquad (9)$$

which means that $\Psi_L(\theta)$ is a strictly decreasing function in θ. The steepest slope can be determined by analyzing the curvature

$$\Psi''_L(\theta) = -(L-1)\theta^{L-2}e^{-\theta} + e^{-\theta}\theta^{L-1}.$$

Equating this to zero will yield $\theta^* = L - 1$. At this critical point and assume $L \gg 1$, a new result using Stirling's formula can be shown[1]:

$$\Psi'_L(\theta^*) \approx -\frac{1}{\sqrt{2\pi\theta^*}} \exp\left\{-\frac{(\theta^* - (L-1))^2}{2\theta^*}\right\}. \qquad (10)$$

Therefore, $\Psi'_L(\theta^*) = -\frac{1}{\sqrt{2\pi(L-1)}}$. Hence, the slope of the incomplete Gamma function reduces as L increases.

Remark 1: Most papers in the image sensor's literature plot curves with respect to $\log_{10}\theta$ instead of θ, like the one shown in Figure 2. The x-axis compression caused by $\log_{10}\theta$ will make the transient of the incomplete Gamma function to appear steeper. The reason is that for any function $f(\theta)$, the slope in the $\log_{10}\theta$ space is determined by $\frac{d}{d\log_{10}\theta}f(\theta) = \theta f'(\theta) \cdot \log 10$. So for large θ, the slope appears steeper.

D. Delta Method

A mathematical tool that will become useful later in the paper is the *Delta Method* in statistics. It approximates the variance when a random variable undergoes a nonlinear transformation.

Lemma 1 (Delta Method, [12]): Consider a sequence of independently and identically distributed (i.i.d.) random variables $X_1, \ldots, X_N$ with a common mean $\mathbb{E}[X_1] = \mu$. Define $\overline{X} = (1/N)\sum_{n=1}^N X_n$ be the sample average, and assume that $\sqrt{N}(\overline{X} - \mu) \xrightarrow{d} \text{Gaussian}(0, \tau^2)$, where τ^2 is the standard deviation of a Gaussian random variable, and $\xrightarrow{d}$ denotes convergence in distribution. Suppose that there is a continuously differentiable function f such that $f'(\mu)$ exists and is not zero. Then $\sqrt{N}[f(\overline{X}) - f(\mu)] \xrightarrow{d} \text{Gaussian}(0, \tau^2(f'(\mu))^2)$. In other words,

$$\mathbb{E}[(f(\overline{X}) - f(\mu))^2] \approx [f'(\mu)]^2 \text{Var}[\overline{X}]. \qquad (11)$$

[1]The proof of this result is given in the Appendix.

Proof: The complete proof can be found in [12, Theorem 2.5.2]. The two key arguments in the proof are (1) $\overline{X}$ converges in distribution to μ due to the Central Limit Theorem; (2) Taylor approximation gives

$$f(\overline{X}) \approx f(\mu) + f'(\mu)(\overline{X} - \mu) + o_p(\overline{X} - \mu),$$

where o_p denotes the little-o_p notation [12, Definition 2.1.3] and hence

$$\sqrt{N}\left[f(\overline{X}) - f(\mu)\right] \approx \sqrt{N}f'(\mu)(\overline{X} - \mu) + o_p(\sqrt{N}(\overline{X} - \mu)).$$

Taking squares on both sides gives (11). $\qquad \square$

III. SNR: A STATISTICAL DEFINITION

A. Defining the SNR

When defining the SNR, it is important to clarify the notion of signal and noise. In most of the imaging problems, the underlying signal is the scene exposure θ. The signal θ defines a probability distribution $p_Y(y; \theta)$ from which the i.i.d. samples $Y_1, \ldots, Y_N$ are drawn.

Reconstruction of the signal θ from $Y_1, \ldots, Y_N$ is based on an *estimator* $\widehat{\theta}$. An estimator can be any mapping that maps $\mathbf{Y} = [Y_1, \ldots, Y_N]$ to $\widehat{\theta}(\mathbf{Y})$. However, if θ is a fixed scalar, most sensors will produce an estimate based on the sample average $\overline{Y} = (1/N)\sum_{n=1}^N Y_n$ because the on-chip processing today is largely limited to simple operations such as addition. In this case, the estimator can be written as $\widehat{\theta}(\overline{Y})$, which can also be interpreted as an estimator based on the sufficient statistics.

The noise term in the SNR is the deviation between the estimator $\widehat{\theta}(\mathbf{Y})$ and the true parameter θ. Since the estimator $\widehat{\theta}(\mathbf{Y})$ is random, the noise power is the expectation

$$\text{noise power} = \mathbb{E}[(\widehat{\theta}(\mathbf{Y}) - \theta)^2], \qquad (12)$$

which is also the mean squared error.

Definition 2 (SNR, formal definition): Let $\mathbf{Y} = [Y_1, \ldots, Y_N]$ be i.i.d. random variables drawn from the distribution $p_Y(y; \theta)$. Construct an estimator $\widehat{\theta}(\mathbf{Y})$. Then the signal-to-noise ratio (SNR) is

$$\text{SNR}(\theta) \stackrel{\text{def}}{=} \frac{\theta}{\sqrt{\mathbb{E}[(\widehat{\theta}(\mathbf{Y}) - \theta)^2]}}. \qquad (13)$$

The above definition is not a new invention. When Yang et al. presented the analysis of the quanta image sensor in 2011 [7], this definition was already used to derive the Cramer-Rao lower bound. Subsequent papers such as [5] and [13] by Chan and colleagues also rely on this definition. A version for single-photon avalanche diode (SPAD) is presented by Gupta and colleagues [14].

To elaborate on this formal definition of the SNR, consider the two examples below.

Example 1 (Poisson): Let $Y_1, \ldots, Y_N \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\theta)$, and consider the maximum-likelihood estimator $\widehat{\theta}(\overline{Y}) = \overline{Y}$. (The derivation of the ML estimator for a Poisson random variable is skipped for brevity.) Since $\widehat{\theta}(\overline{Y}) = \overline{Y}$, it follows that

$$\mathbb{E}[(\widehat{\theta}(\overline{Y}) - \theta)^2] = \mathbb{E}[(\overline{Y} - \theta)^2] = \frac{\theta}{N},$$

where the second equality holds because the variance of a Poisson random variable is θ. Therefore, $\text{SNR}(\theta) = \sqrt{N} \cdot \sqrt{\theta}$, which is consistent with (5) when $\sigma_{\text{read}} = 0$. □

Example 2 (Poisson + Gaussian): Let $Y_1, \ldots, Y_N \stackrel{\text{i.i.d.}}{\sim}$ Poisson(θ) + Gaussian$(0, \sigma_{\text{read}}^2)$. Consider the ML estimator $\widehat{\theta}(\overline{Y}) = \overline{Y}$. It then follows that

$$\mathbb{E}[(\widehat{\theta}(\overline{Y}) - \theta)^2] = \mathbb{E}[(\overline{Y} - \theta)^2] = \frac{1}{N}\left(\theta + \sigma_{\text{read}}^2\right),$$

where the second equality holds because the variance of a Poisson-Gaussian is the sum of the two variances. Therefore, $\text{SNR}(\theta) = \sqrt{N} \cdot \theta / \sqrt{\theta + \sigma_{\text{read}}^2}$. This result is consistent with (5) for a general σ_{read}. □

B. Mean Invariance

The formal definition of the SNR is general for *any* estimator. However, the statistical noise model for an actual image sensor can be complicated. In fact, except for a few special occasions where the maximum-likelihood (ML) estimator can be expressed in a closed form, in most other situations the ML estimator cannot be obtained in a closed form. In this case, a more convenient way is to define the estimator from the mean.

Definition 3 (Mean invariant estimator): Let $Y_1, \ldots, Y_N$ be i.i.d. random variables drawn from the distribution $p_Y(y; \theta)$. Let $\mu(\theta) = \mathbb{E}[Y] = \int y \cdot p_Y(y; \theta)\,dy$ be the mean of Y. An estimator $\widehat{\theta}(\boldsymbol{Y})$ is mean invariant if

$$\mu(\widehat{\theta}(\boldsymbol{Y})) = \overline{Y}. \tag{14}$$

If μ^{-1} exists, the mean invariant estimator is $\widehat{\theta}(\boldsymbol{Y}) = \mu^{-1}(\overline{Y})$.

The following two examples shows that many estimators are mean invariant.

Example 3: Let $Y_n \stackrel{\text{i.i.d.}}{\sim}$ Gaussian$(\theta, \sigma_{\text{read}}^2)$ for $n = 1, \ldots, N$, where θ is the unknown parameter. It can be shown that the maximum-likelihood (ML) estimator is

$$\widehat{\theta}_{\text{ML}}(\overline{Y}) = \underset{\theta}{\text{argmax}} \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{\text{read}}^2}} \exp\left\{-\frac{(Y_n - \theta)^2}{2\sigma_{\text{read}}^2}\right\} = \overline{Y}.$$

Notice that if Y is Gaussian, the mean is $\mu(\theta) = \mathbb{E}[Y] = \theta$. Therefore, the mean invariance property holds:

$$\mu(\widehat{\theta}_{\text{ML}}(\overline{Y})) \stackrel{(a)}{=} \widehat{\theta}_{\text{ML}}(\overline{Y}) = \overline{Y}.$$

where (a) is due to the fact that $\mu(\theta) = \theta$. □

Example 4: Let $Y_1, \ldots, Y_N$ be i.i.d. one-bit measurements defined according to (6), with $q = 0.5$ and $\sigma_{\text{read}} = 0$. In this case, Y_n is a Bernoulli random variable such that $Y_n \stackrel{\text{i.i.d.}}{\sim}$ Bernoulli$(1 - e^{-\theta})$. The ML estimator is

$$\widehat{\theta}_{\text{ML}}(\overline{Y}) = \underset{\theta}{\text{argmax}} \prod_{n=1}^{N} (1 - e^{-\theta})^{Y_n}(e^{-\theta})^{1-Y_n}$$
$$= -\log(1 - \overline{Y}).$$

Since the mean of Y is $\mu(\theta) = \mathbb{E}[Y] = 1 - e^{-\theta}$, it follows that

$$\mu(\widehat{\theta}_{\text{ML}}(\overline{Y})) = 1 - e^{-\widehat{\theta}_{\text{ML}}(\overline{Y})} = \overline{Y}.$$

Again, the mean invariance property is satisfied. □

A mean invariant estimator is easy to construct. Even if $p_Y(y; \theta)$ has a complex form, the mean $\mathbb{E}[Y]$ can be obtained through Monte Carlo simulation. Once the mean $\mathbb{E}[Y]$ is determined, the mean invariant estimator $\widehat{\theta}$ can be *constructed* from $\widehat{\theta}(\overline{Y}) = \mu^{-1}(\overline{Y})$, assuming that μ^{-1} exists.

Follow up of Example 3. Let $Y_n \stackrel{\text{i.i.d.}}{\sim}$ Gaussian$(\theta, \sigma_{\text{read}}^2)$ for $n = 1, \ldots, N$ where θ is the unknown parameter. Since the mean is $\mathbb{E}[Y] = \theta$, it follows that $\mu(\theta) = \mathbb{E}[Y] = \theta$. This μ is the identity mapping, and so the inverse mapping is $\mu^{-1}(s) = s$ for any s. Thus, one can define an estimator $\widehat{\theta}(\overline{Y}) = \mu^{-1}(\overline{Y}) = \overline{Y}$. As seen, it is the same as the ML estimator derived in Example 3. Moreover, $\widehat{\theta}(\overline{Y})$ is mean invariant because $\widehat{\theta}$ is constructed in that way. □

Follow up of Example 4. Let $Y_n \stackrel{\text{i.i.d.}}{\sim}$ Bernoulli$(1 - e^{-\theta})$ be the one-bit measurements for $n = 1, \ldots, N$. The mean is $\mu(\theta) = \mathbb{E}[Y] = 1 - e^{-\theta}$, and so the inverse is $\mu^{-1}(s) = -\log(1 - s)$. Therefore, one can define the estimator as $\widehat{\theta}(\overline{Y}) = \mu^{-1}(\overline{Y}) = -\log(1 - \overline{Y})$ The result is identical to Example 4. Furthermore, since the estimator is constructed from the mean invariance property, it has to satisfy the property. □

Based on Example 3 and Example 4, one may conjecture that any ML estimator is also the mean invariant estimator. The observation is correct for any distributions in the exponential family. The proof is given in the Appendix. Outside the exponential family the two can be different. The following is a counter example.

Example 5 (Mean invariant estimator ≠ ML estimator): Consider a truncated Poisson distribution

$$p_Y(y; \theta) = \begin{cases} \frac{\theta^y e^{-\theta}}{y!}, & y < L, \\ 1 - \Psi_L(\theta), & y \geq L. \end{cases}$$

Using (23) (to be proved in the next section), the mean is $\mathbb{E}[Y] = \theta\Psi_{L-1}(\theta) + L(1 - \Psi_L(\theta))$. Let $\mu(\theta) = \theta\Psi_{L-1}(\theta) + L(1 - \Psi_L(\theta))$. So, the mean invariant estimator can be defined as

$$\widehat{\theta}(\boldsymbol{Y}) = \mu^{-1}(\overline{Y}),$$

which is a function of $\overline{Y} = (1/N)\sum_{n=1}^{N} Y_n$.

Now consider the ML estimator. The ML estimator is

$$\widehat{\theta}_{\text{ML}}(\boldsymbol{Y}) = \underset{\theta}{\text{argmax}} \quad \frac{1}{N}\sum_{n=1}^{N}\left[(Y_n\log\theta - \theta)\cdot\underbrace{\mathbb{I}\{Y_n < L\}}_{Z_n}\right.$$
$$\left. + \log(1 - \Psi_L(\theta))\cdot\underbrace{\mathbb{I}\{Y_n \geq L\}}_{1 - Z_n}\right], \tag{15}$$

where $Z_n = \mathbb{I}\{Y_n < L\}$ is the indicator function that returns 1 if $Y_n < L$ or 0 if otherwise. Taking derivative and setting it to zero implies that $\widehat{\theta}_{\text{ML}}$ must satisfy the equation

$$\frac{1}{\widehat{\theta}_{\text{ML}}}\left(\frac{1}{N}\sum_{n=1}^{N}Y_nZ_n\right) - \frac{1}{N}\sum_{n=1}^{N}Z_n$$
$$= \left(1 - \frac{1}{N}\sum_{n=1}^{N}Z_n\right)\cdot\frac{\Psi_L'(\widehat{\theta}_{\text{ML}})}{1 - \Psi_L(\widehat{\theta}_{\text{ML}})}. \tag{16}$$

Therefore, the ML estimator $\widehat{\theta}_{\text{ML}}(\boldsymbol{Y})$ must be a function of $(1/N)\sum_{n=1}^{N}Y_nZ_n$ and $(1/N)\sum_{n=1}^{N}Z_n$, not $\overline{Y}$. □

While the ML estimator and the mean invariant estimator are generally not the same, they are asymptotically equivalent. As $N \to \infty$, the consistency of the ML estimator implies that $\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{Y}) \xrightarrow{p} \theta$ [12]. On the other hand, the law of large numbers implies that $\overline{Y} \xrightarrow{p} \mathbb{E}[Y] = \mu(\theta)$. So, a mean invariant estimator $\widehat{\theta}(\boldsymbol{Y}) = \mu^{-1}(\overline{Y}) \xrightarrow{p} \mu^{-1}(\mu(\theta)) = \theta$. Therefore, as $N \to \infty$, both $\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{Y})$ and $\widehat{\theta}(\boldsymbol{Y})$ will converge in probability to the true parameter θ and hence they are equivalent asymptotically.

Mean invariance is also not the same as the invariance principle of the ML estimator. The invariance principle of the ML estimator says that if there is a monotonic mapping h that maps θ to $h(\theta)$, then $h(\widehat{\theta}_{\mathrm{ML}})$ will be the ML estimator of $h(\theta)$. The following example shows a case where mean invariance is different from the invariance of ML.

Example 6: Consider $X_n \overset{\text{i.i.d.}}{\sim}$ Bernoulli(θ) and $Y_n \overset{\text{i.i.d.}}{\sim}$ Bernoulli(e^θ) for $n = 1, \dots, N$. The ML estimator of θ using $X_1, \dots, X_N$ is
$$\widehat{\theta}_{\mathrm{ML}}(\overline{X}) = \overline{X}.$$

According to the invariance of the ML estimator, a monotonic mapping $h_1(\cdot) = e^{(\cdot)}$ will ensure that $\widehat{\theta}_1 \overset{\text{def}}{=} h_1(\widehat{\theta}_{\mathrm{ML}}) = e^{\overline{X}}$ is the ML estimator of e^θ. By the same principle, if there is a different monotonic mapping $\widehat{\theta}_2 \overset{\text{def}}{=} h_2(\cdot) = \log(\cdot)$, it holds that $h_2(\widehat{\theta}_{\mathrm{ML}}) = \log(\overline{X})$ is the ML estimator of $\log(\theta)$.

Now consider the mean $\mu(\theta) \overset{\text{def}}{=} \mathbb{E}[Y] = e^\theta$. The invariance principle says nothing about whether $\mu(\widehat{\theta}_1(\overline{Y})) = \overline{Y}$ or $\mu(\widehat{\theta}_2(\overline{Y})) = \overline{Y}$. In fact, $\widehat{\theta}_1$ does not satisfy the mean invariance property because $\mu(\widehat{\theta}_1(\overline{Y})) = e^{\widehat{\theta}_1(\overline{Y})} = e^{e^{\overline{Y}}} \neq \overline{Y}$. However, $\widehat{\theta}_2$ satisfies the mean invariance property because $\mu(\widehat{\theta}_2(\overline{Y})) = e^{\log(\overline{Y})} = \overline{Y}$. Therefore, the invariance principle of the ML estimator is completely different from the mean invariance property. $\square$

To summarize, the mean invariance is a property that specifically focuses on whether the mean $\mathbb{E}[Y]$ can be nonlinearly mapped to recover the true parameter θ. This is the property required for the exposure-referred SNR. Whether the estimator is the ML estimator is not of concern.

In the statistics literature, the mean invariance property presented in this paper is related to the *link function* for the generalized linear models. Specifically, the mapping $\mu(\theta) \overset{\text{def}}{=} \mathbb{E}[Y]$ from the true parameter θ to the mean μ is known as the link function, and the inverse mapping μ^{-1} is known as the response function. Readers interested in details of this topic can consult [15].

C. Exposure-referred SNR

With all the mathematical tools ready, the exposure-referred SNR can now be formally derived.

Theorem 1: Let $Y_1, \dots, Y_N$ be i.i.d. random variables drawn from the probability density function $p_Y(y; \theta)$. Define $\overline{Y} = (1/N) \sum_{n=1}^{N} Y_n$. Let $\mu(\theta) \overset{\text{def}}{=} \mathbb{E}[Y]$ and assume that μ^{-1} exists. Let $\widehat{\theta}(\boldsymbol{Y})$ be the mean invariant estimator such that $\widehat{\theta}(\boldsymbol{Y}) = \mu^{-1}(\overline{Y})$. Then the SNR defined in (13) is related to $\mathrm{SNR}_{\mathrm{exp}}$ as
$$\mathrm{SNR}(\theta) \approx \mathrm{SNR}_{\mathrm{exp}}(\theta) = \frac{\theta}{\sqrt{\mathrm{Var}[\overline{Y}]}} \cdot \frac{d\mu}{d\theta}. \tag{17}$$

Proof: By the Delta Method, the mean squared error can be approximated by
$$\mathbb{E}\left[\left(\widehat{\theta}(\overline{Y}) - \theta\right)^2\right] = \mathbb{E}\left[\left(\widehat{\theta}(\overline{Y}) - \widehat{\theta}(\mu(\theta))\right)^2\right]$$
$$\approx \left[\widehat{\theta}'(\mu(\theta))\right]^2 \mathrm{Var}[\overline{Y}],$$

where the derivative is taken with respect to $\mu(\theta)$.

Since $\widehat{\theta}(\mu(\theta)) = \mu^{-1}(\mu(\theta)) = \theta$, it follows that $\frac{d\widehat{\theta}(\mu)}{d\mu} = \frac{d\theta}{d\mu}$. So,
$$\mathbb{E}\left[\left(\widehat{\theta}(\overline{Y}) - \theta\right)^2\right] = \left[\frac{d\theta}{d\mu}\right]^2 \mathrm{Var}[\overline{Y}]. \tag{18}$$

Using the fact that $\frac{d\theta}{d\mu} = 1 / \frac{d\mu}{d\theta}$, the SNR can be written as
$$\mathrm{SNR}(\theta) = \frac{\theta}{\sqrt{\mathbb{E}\left[\left(\widehat{\theta}(\overline{Y}) - \theta\right)^2\right]}} \approx \underbrace{\frac{\theta}{\sqrt{\mathrm{Var}[\overline{Y}]}} \cdot \frac{d\mu}{d\theta}}_{\mathrm{SNR}_{\mathrm{exp}}(\theta)},$$
which completes the proof. $\square$

Corollary 1: Under the same conditions listed in Theorem 1, the exposure-referred SNR is related to the output-referred SNR as
$$\mathrm{SNR}_{\mathrm{exp}}(\theta) = \mathrm{SNR}_{\mathrm{out}}(\theta) \cdot \frac{\theta}{\mu} \cdot \frac{d\mu}{d\theta}. \tag{19}$$

Proof: The proof follows from the substitution
$$\mathrm{SNR}_{\mathrm{exp}}(\theta) = \frac{\theta}{\sqrt{\mathrm{Var}[\overline{Y}]}} \cdot \frac{d\mu}{d\theta} = \underbrace{\frac{\mu}{\sqrt{\mathrm{Var}[\overline{Y}]}}}_{=\mathrm{SNR}_{\mathrm{out}}(\theta)} \cdot \frac{\theta}{\mu} \cdot \frac{d\mu}{d\theta}.$$

This completes the proof. $\square$

As one can see from (18), the derivative $d\mu/d\theta$ is added because of the Delta Method. It is the first-order approximation of a nonlinear mapping from the output to the input. Using the argument of Elgendy and Chan [5], this first-order derivative can be regarded as a transfer function relating the output $\mathbb{E}[\overline{Y}]$ to the input θ. If the input-output has a linear relationship $\mathbb{E}[\overline{Y}] = \theta$, which is the case of a CIS with a large full-well capacity, then the derivative is $d\mu/d\theta = 1$ and so $\mathrm{SNR}_{\mathrm{out}}(\theta) = \mathrm{SNR}_{\mathrm{exp}}(\theta)$.

D. Illustrating the SNR via one-bit QIS

To elaborate on the difference between $\mathrm{SNR}_{\mathrm{exp}}(\theta)$ and $\mathrm{SNR}_{\mathrm{out}}(\theta)$, it would be instructive to consider the statistics of a one-bit quanta image sensor. Let $X_1, \dots, X_N \overset{\text{i.i.d.}}{\sim}$ Poisson(θ) and let Y_n be the random variables defined in (6).

First, consider the case where $q = 1$. Since $Y_1, \dots, Y_N \overset{\text{i.i.d.}}{\sim}$ Bernoulli$(1 - e^{-\theta})$, the mean is $\mathbb{E}[Y] = 1 - e^{-\theta}$. Define the mean as $\mu(\theta) = \mathbb{E}[Y] = 1 - e^{-\theta}$. As shown in Example 4, the maximum-likelihood estimate of θ is $\widehat{\theta}(\overline{Y}) = -\log(1 - \overline{Y})$ and it satisfies the mean invariance property. The derivative $d\mu/d\theta$ is
$$\frac{d\mu}{d\theta} = \frac{d}{d\theta}\left[1 - e^{-\theta}\right] = e^{-\theta}.$$

Substituting into Theorem 1, it can be shown that

$$\text{SNR}_{\text{exp}}(\theta) = \frac{\theta}{\sqrt{\text{Var}[\overline{Y}]}} \cdot \frac{d\mu}{d\theta} = \sqrt{N} \cdot \theta \cdot \sqrt{\frac{e^{-\theta}}{1 - e^{-\theta}}}.$$

For cases where $q > 1$, one can use the incomplete Gamma function so that

$$p_Y(y; \theta) = \begin{cases} 1 - \Psi_q(\theta), & y = 1, \\ \Psi_q(\theta), & y = 0. \end{cases}$$

It then follows that $\mathbb{E}[Y] = 1 - \Psi_q(\theta)$ and the estimator can be chosen such that $\widehat{\theta}(\overline{Y}) = \Psi_q^{-1}(1 - \overline{Y})$. The mean invariance property is therefore validated. The derivative $d\mu/d\theta$ is

$$\frac{d\mu}{d\theta} = \frac{d}{d\theta}(1 - \Psi_q(\theta)) = \frac{\theta^{q-1} e^{-\theta}}{(q-1)!}.$$

Hence, the SNR is

$$\text{SNR}_{\text{exp}}(\theta) = \frac{\sqrt{N} \cdot \theta}{\sqrt{\Psi_q(\theta)(1 - \Psi_q(\theta))}} \cdot \frac{\theta^{q-1} e^{-\theta}}{(q-1)!}, \quad (20)$$

of which the visualization is shown in Figure 3. This result is consistent with the one shown by Elgendy and Chan by deriving the Fisher Information [5].
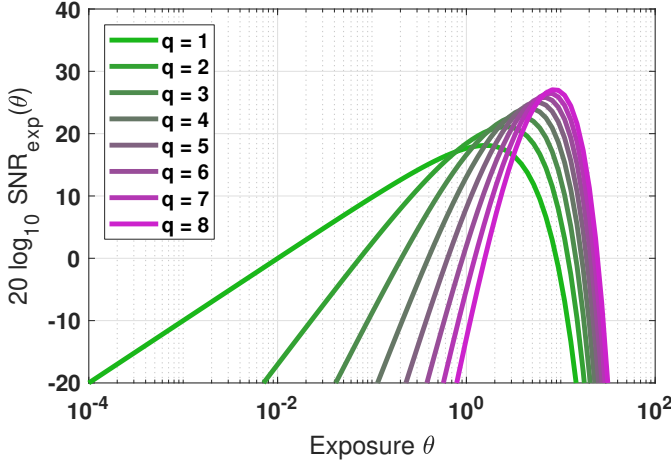

Fig. 3. Exposure-referred SNR for a one-bit quanta image sensor, assuming that $\sigma_{\text{read}} = 0$. As θ goes beyond the threshold q, $\text{SNR}_{\text{exp}}(\theta)$ starts to drop as expected.

Unlike $\text{SNR}_{\text{exp}}(\theta)$, the output-referred SNR goes to infinity when θ grows. For the same one-bit statistics, the output-referred SNR is simply the ratio between $\mathbb{E}[\overline{Y}]$ and $\text{Var}[\overline{Y}]$, which is

$$\text{SNR}_{\text{out}}(\theta) = \frac{\mathbb{E}[\overline{Y}]}{\sqrt{\text{Var}[\overline{Y}]}} = \sqrt{N} \cdot \sqrt{\frac{1 - \Psi_q(\theta)}{\Psi_q(\theta)}}. \quad (21)$$

As shown in Figure 4, $\text{SNR}_{\text{out}}(\theta)$ grows indefinitely as θ grows. This does not reflect the reality because when θ grows beyond the threshold q, the measurements $\{Y_n \mid n = 1, \ldots, N\}$ will have more one's. The signal degrades and hence eventually the SNR drops to zero.
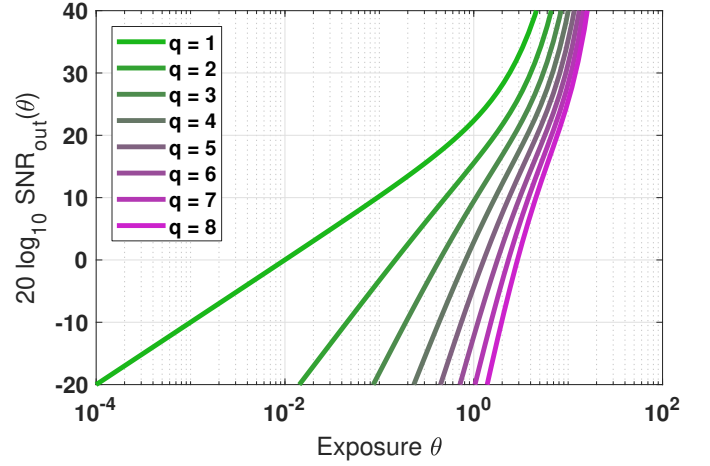

Fig. 4. Output-referred SNR for a one-bit quanta image sensor, assuming that $\sigma_{\text{read}} = 0$. As θ goes beyond the threshold q, $\text{SNR}_{\text{out}}(\theta)$ continues to grow because of the inability of $\text{SNR}_{\text{out}}(\theta)$ to handle pixel saturation.

IV. $\text{SNR}_{\text{EXP}}(\theta)$ FOR FINITE FULL-WELL CAPACITY

The subject of this paper is to understand the exposure-referred SNR for digital (CCD and CMOS) image sensors with a finite full-well capacity L. In particular, the goal is to understand the situation when L is small, e.g., a few bits. This section presents the main result for such a scenario.

A. $SNR_{exp}(\theta)$ for Truncated Poisson

To make the analysis tractable, the derivation in this section will be focusing on a truncated Poisson distribution assuming $\sigma_{\text{read}} = 0$. Extension to the more complex noise model will be analyzed later.

Theorem 2 (SNR$_{exp}(\theta)$ for truncated Poisson): Let $Y_1, \ldots, Y_N$ be i.i.d. random variables following the truncated Poisson statistics defined in (4) where $X_n \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\theta)$ for $n = 1, \ldots, N$. Let $\widehat{\theta}(\overline{Y})$ be an estimator satisfying the mean invariance property. Then the exposure-referred SNR is

$$\text{SNR}_{\text{exp}}(\theta) = \sqrt{N} \cdot \frac{\theta}{\sqrt{\text{Var}[Y_n]}} \cdot \frac{d\mu}{d\theta}, \quad (22)$$

where

$$\mathbb{E}[Y] = \theta\Psi_{L-1}(\theta) + L(1 - \Psi_L(\theta)) \overset{\text{def}}{=} \mu,$$
$$\text{Var}[Y] = \theta^2\Psi_{L-2}(\theta) + \theta\Psi_{L-1}(\theta) + L^2(1 - \Psi_L(\theta)) - \mu^2,$$
$$\frac{d\mu}{d\theta} = \theta\Psi'_{L-1}(\theta) + \Psi_{L-1}(\theta) - L\Psi'_L(\theta). \quad (23)$$

Proof: The proof is presented in the Appendix. □

To illustrate the predicted $\text{SNR}_{\text{exp}}(\theta)$ as a function of θ, Figure 5 shows several curves evaluated at different full-well capacity L. As is consistent with the one-bit QIS example shown in Section III.D, the exposure-referred SNR for a truncated Poisson random variable also demonstrates a drop in $\text{SNR}_{\text{exp}}(\theta)$ after the pixel saturates. What is more interesting is that as L increases, $\text{SNR}_{\text{exp}}(\theta)$ becomes a straight line in the log-log plot with a sharp decay after saturation.

The rapid drop after the saturation is attributed to two reasons. First, as explained in the remark in Section II.C, the log-log plot has a compression of the x-axis so that the slope
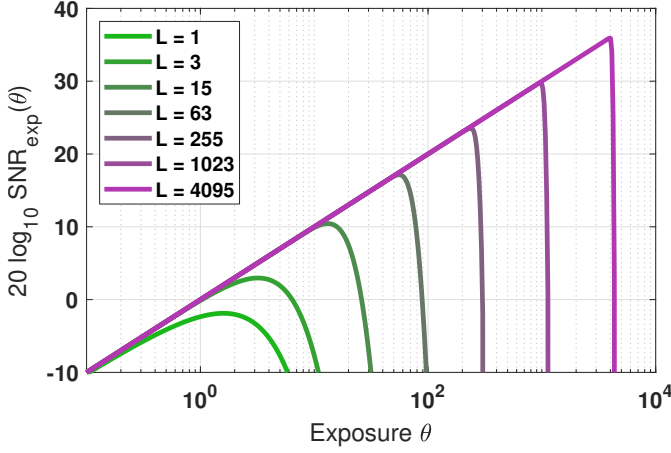
Fig. 5. Exposure-referred SNR for a digital image sensor with a full-well capacity of L electrons.

is amplified with θ. If one plots the x-axis in the linear scale (instead of the log scale), the sharp cutoff will appear in a smoother transition. However, since in practice the exposure is always shown in the log scale, what is being shown in Figure 5 is valid. The second reason for the drop after the saturation is due to the limiting behavior of the incomplete Gamma function. As L increases, the incomplete Gamma function in the log-log plot will have an increasingly sharp transient as shown in Figure 2. This will be shown theoretically in the next subsection.

B. Limiting Case

Figure 5 shows that as the full-well capacity L increases, $\text{SNR}_{\text{exp}}(\theta)$ becomes more linear in the log-log plot. Such a behavior can be theoretically derived by analyzing the limiting cases of the incomplete Gamma function. The log-log plot requires the x-axis to be mapped from θ to $\log_{10}\theta$. In this case, define $\phi = \log_{10}\theta$ (so that $\theta = 10^\phi$) and it can be shown that

$$\Psi_L(10^\phi) \approx \begin{cases} 1, & \phi \leq \log_{10} L, \\ 0, & \phi > \log_{10} L, \end{cases} \quad (24)$$

assuming that $L \gg 1$.

Corollary 2: Consider the same conditions as in Theorem 2 but with $L \gg 1$. Let $\Psi = \Psi_L(\theta)$. Under the limiting assumption of $\Psi_L(\theta)$ described in (24), it holds that

$$20\log_{10}\text{SNR}_{\text{exp}}(10^\phi) = \begin{cases} 10\phi, & \phi \leq \log_{10} L, \\ -\infty, & \phi > \log_{10} L. \end{cases} \quad (25)$$

Proof: See Appendix. □

The implication of the corollary is that as L increases, plotting $\text{SNR}_{\text{exp}}(\theta)$ in the log-log plot will give a linear response followed by an abrupt transition. This is exactly what is happening in the output-referred SNR. Therefore, Theorem 2 is a generalized version of the output-referred SNR curves reported in the literature. For practical algorithms such as those for high dynamic range imaging, (25) is very common, for example used in [16].

V. MONTE CARLO SIMULATION

So far, the theoretical derivations have been focusing on the Poisson distribution only. Read noise, dark current, quantization error, and other sources of noise have not been considered. When including these factors, seeking an analytic expression would be significantly more challenging. A more reasonable approach is to resort to numerical schemes to estimate the SNR approximately.

A. General Principle

In general, the measurement Y generated by an image sensor is the result of a sequence of optical-electronic operations such as

$$Y = \text{clip}\left\{\text{round}\left\{\text{Poisson}(\theta + \theta_{\text{dark}}) + \text{Gauss}(0, \sigma_{\text{read}}^2)\right\}\right\}, \quad (26)$$

where θ_{dark} denotes the dark current, "round" denotes the analog-to-digital (A/D) conversion, and "clip" denotes the saturation due to a finite full-well capacity. Assuming a sufficiently large full-well capacity L, the output-referred SNR is given by [16], [17]

$$\text{SNR}_{\text{out}}(\theta) = \begin{cases} \frac{\theta}{\sqrt{\theta + \theta_{\text{dark}} + \sigma_{\text{read}}^2}}, & \theta < L, \\ 0, & \theta \geq L, \end{cases} \quad (27)$$

where L is the full-well capacity.

To compute $\text{SNR}_{\text{exp}}(\theta)$, the numerical approach is to draw samples from the the distribution defined by the forward model:

$$Y_m = \text{forward model}\left(\theta \mid \theta_{\text{dark}}, \sigma_{\text{read}}, L\right), \quad (28)$$

for $m = 1, \ldots, M$, where M denotes the number of Monte Carlo samples. As stated in (28), the mth sample Y_m is a function of the underlying signal θ, along with other model parameters. Do not confuse M with the number of i.i.d. measurements N used in the previous subsections when defining the average $\overline{Y}$.

The Monte Carlo sampling scheme goes as follow. Consider $N = 1$. For every θ, the sample average is an estimate of $\mathbb{E}[Y]$ and the sample variance is an estimate of $\text{Var}[Y]$:

$$\widehat{\mu}(\theta) = \frac{1}{M}\sum_{m=1}^{M} Y_m, \quad \text{and} \quad \widehat{\sigma}^2(\theta) = \frac{1}{M}\sum_{m=1}^{M}(Y_m - \widehat{\mu})^2.$$

Once $\widehat{\mu}(\theta)$ has been determined for every θ, the derivative $d\mu/d\theta$ can be approximated by

$$\frac{d\widehat{\mu}}{d\theta} = \frac{\widehat{\mu}(\theta_{k+1}) - \widehat{\mu}(\theta_k)}{\theta_{k+1} - \theta_k},$$

where $\{\theta_k \mid \theta_k < \theta_{k+1}, \ k = 1, \ldots, K\}$ is the discrete set of exposures used to evaluate the mean and variance. Consequently, if there are N i.i.d. samples, $\text{SNR}_{\text{exp}}(\theta)$ can be approximately estimated by

$$\widehat{\text{SNR}}_{\text{exp}}(\theta) = \sqrt{N} \cdot \frac{\theta}{\widehat{\sigma}} \cdot \frac{d\widehat{\mu}}{d\theta}. \quad (29)$$

B. *Visualizing the Impacts of θ_{dark} and σ_{read}*

With the Monte Carlo simulation technique, complex forward models can be visualized. Consider the following two demonstrations.

Example 7 (Influence of Read Noise): The first scenario considers a fixed dark current, full-well capacity, and A/D converter, but a varying read noise level. Let $\theta_{dark} = 0.016$ (which is consistent with the quanta image sensor [18]), a full-well capacity of $L = 15$ electrons, and 4-bit A/D converter. The read noise level σ_{read} varies from 0 to 4.5 with a step interval of 0.5. By using $M = 5 \times 10^6$ Monte Carlo samples, the numerically simulated $\text{SNR}_{exp}(\theta)$ is plotted in Figure 6.
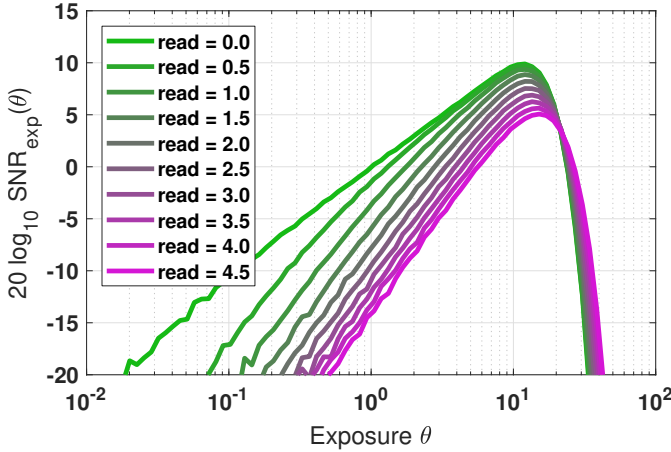

Fig. 6. Exposure-referred SNR for a digital image sensor by considering different levels of read noise.

Increasing the read noise leads to a reduced SNR for all θ before saturation. After saturation, the read noise will occasionally move a saturated measurement back to an unsaturated state because the Gaussian noise can take a negative value. See that the purple curves on the right-hand side of the plot are higher than the green curves. Therefore, for large θ, there is a minor but noticeable gain in SNR, especially when the read noise is high. This is not necessarily a better outcome, because the increased SNR at larger θ comes at the cost of significantly lower SNR in low light where the θ is small.

Remark: The small fluctuation towards the tail in Figure 6 is due to the randomness in the Monte Carlo simulation. As M goes to infinity, the random estimate will approach the expectation by the law of large number.   □

Example 8 (Influence of Dark Current): The second scenario considers a fixed read noise, full-well capacity, and A/D converter, but a varying dark current. To be consistent with the literature, the dark current θ_{dark} is assumed to vary from 0 to 0.45 with a step interval of 0.05. The read noise level is fixed at 0.2 based on [18]. The full-well capacity is $L = 15$ electrons, and a 4-bit A/D converter is used. Same as Example 7, $M = 5 \times 10^6$ Monte Carlo samples are used to numerically generate the $\text{SNR}_{exp}(\theta)$ plot in Figure 7.

Unlike Example 7 where the read noise has a substantial influence to the SNR, an increased dark current will only show its impact for small θ. This should not be a surprise because when the true signal θ is strong, the influence of θ_{dark} will be negligible considering the small magnitude it usually has. For

small θ, the impact of θ_{dark} is more prominent. A smaller dark current indeed leads to a higher SNR as expected.   □
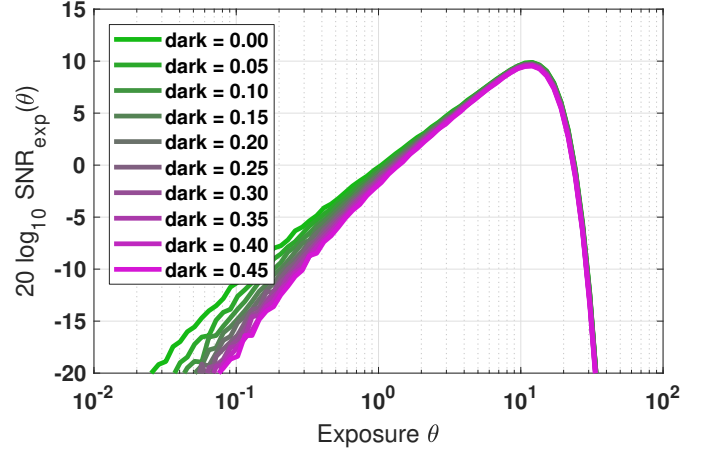

Fig. 7. Exposure-referred SNR for a digital image sensor by considering different levels of dark current. The small fluctuation towards the tail on the left-hand side is due to randomness in the Monte Carlo simulation.

The utility of the Monte Carlo simulation is that it bypasses the complication of seeking for an analytic expression of $\text{SNR}_{exp}(\theta)$. To account for even more difficult modelings such as the pixel response non-uniformity, $1/f$ noise, conversion gain, and exposure time, etc, one just needs to modify the forward image formation model. For extreme cases such as very small θ where the random fluctuation is significant, one easy fix is to approximate $\text{SNR}_{exp}(\theta)$ by $\text{SNR}_{out}(\theta)$ using (27). This approximation is reasonably accurate for small θ that is sufficiently far away from the saturation cutoff.

VI. Utilities

After elaborating on the details of the exposure-referred SNR, readers may ask: what are the utilities of this SNR? The answer is simple. As a performance metric of an image sensor, the primary utility of the exposure-referred SNR is to describe how well an image sensor performs. Because of this primary goal, three points should be noted:

- The exposure-referred SNR is a generalized version of the output-referred SNR. The latter is a special case of the former when the bit-depth L is large as shown in Figure 8. The output-referred SNR and the exposure-referred SNR are very similar for large L, whereas, for a small bit-depth, the output-referred SNR cannot capture the phenomenon when the exposure goes beyond the full-well capacity.

- Because of the first point, any subsequent low bit-depth sensors applications need to use the exposure-referred SNR. Using the output-referred SNR will lead to sub-optimal performance, and this will be illustrated in Application 1 on high-dynamic range imaging.

- Again, because of the first point, the exposure-referred SNR can be used as the objective function to tune the parameters of the one-bit and few-bit image sensors. Since closed-form expressions of the SNR are available for some cases, the optimal parameters can also be
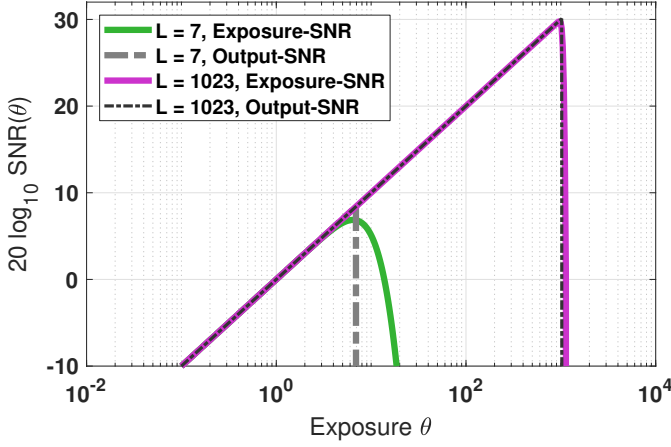
Fig. 8. The output-referred SNR is a special case of the exposure-referred SNR when the full-well capacity L is large. This figure plots the SNRs of a truncated Poisson random variable. For the output-referred SNR, the curve is generated by $\text{SNR}_{\text{out}}(\theta) = \sqrt{\theta}$ for $\theta \leq L$ and $\text{SNR}_{\text{out}}(\theta) = 0$ for $\theta > L$.

expressed in closed-form. Application 2 on the threshold design will illustrate this point.

A. High Dynamic Range Imaging Using a 3-bit Sensor

The majority of the high dynamic range (HDR) image reconstruction algorithms today are designed for digital image sensors with a large full-well capacity, and so $\text{SNR}_{\text{out}}(\theta)$ is adequate. However, for image sensors with a small full-well capacity, e.g., $L = 7$, a fusion algorithm based on $\text{SNR}_{\text{exp}}(\theta)$ will produce a better image.

To illustrate the impact of $\text{SNR}_{\text{exp}}(\theta)$ in HDR imaging, consider the problem of reconstructing one HDR image (a single pixel) $\widehat{\theta}$ from M exposure brackets $\boldsymbol{Y}^1, \ldots, \boldsymbol{Y}^M$. Each of these exposure brackets $\boldsymbol{Y}^m = [Y_1^m, \ldots, Y_N^m]$ has N frames. The formation of each frame follows the equation

$$Y_n^m \overset{\text{i.i.d.}}{\sim} \begin{cases} \text{Poisson}(\tau_m\theta), & \tau_m\theta < L, \\ L, & \tau_m\theta \geq L, \end{cases}$$

for $n = 1, \ldots, N$, where τ_m is the mth integration time. To reconstruct the mth low-dynamic range (LDR) image, one can substitute the sample average $\overline{Y}^m = (1/N)\sum_{n=1}^N Y_n^m$ into the inverse mapping of the mean μ^{-1}. Then the HDR image is fused by a linear combination of the estimates [2]

$$\widehat{\theta} = \sum_{m=1}^M w_m \frac{\mu^{-1}(\overline{Y}^m)}{\tau_m}, \tag{30}$$

where $\{w_1, \ldots, w_M\}$ is a set of linear combination weights. The development of the idea can be traced back to Mann and Picard [19], Debevec and Malik [20], among other works [16], [21], [22]. For a conventional CIS with a large full-well capacity, μ^{-1} is an identity mapping and so the reconstruction is simplified to $\widehat{\theta} = \sum_{m=1}^M \frac{w_m}{\tau_m}\overline{Y}^m$. For sensors with a small full-well capacity, the nonlinearity of the mean function

[2]The problem here does not assume any motion so that the theoretically optimal solution can be analytically derived. Handling motion remains an open problem in computer vision, although significant progress has been made over the past decade.

should be taken into account. This reconstruction scheme is elementary and predates all the deep learning methods.

As previously proved by Gnanasambandam and Chan in [13], the optimal weight w_m is

$$w_m = \frac{\text{SNR}_m^2(\theta)}{\sum_{m=1}^M \text{SNR}_m^2(\theta)}, \tag{31}$$

where $\text{SNR}_m(\theta)$ is the SNR of the mth low dynamic range image. If $\text{SNR}_{\text{out}}(\theta)$ is used, the weight will become

$$w_m = \frac{\text{SNR}_{\text{out},m}^2(\theta)}{\sum_{m=1}^M \text{SNR}_{\text{out},m}^2(\theta)} = \frac{(\sqrt{\tau_m\theta \cdot \mathbb{I}_{\{\tau_m\theta<L\}}})^2}{\sum_{m=1}^M (\sqrt{\tau_m\theta \cdot \mathbb{I}_{\{\tau_m\theta<L\}}})^2}$$
$$= \frac{\tau_m \cdot \mathbb{I}_{\{\tau_m\theta<L\}}}{\sum_{m=1}^M \tau_m \cdot \mathbb{I}_{\{\tau_m\theta<L\}}}, \tag{32}$$

where the function $\mathbb{I}_{\{\tau_m\theta<L\}}$ is a binary indicator showing whether the total exposure $\tau_m\theta$ has exceeded the full-well capacity. $\mathbb{I}_{\{\tau_m\theta<L\}} = 1$ if the argument is true and $\mathbb{I}_{\{\tau_m\theta<L\}} = 0$ if the argument is false. If $\text{SNR}_{\text{exp},m}(\theta)$ is used, the optimal weight will become

$$w_m = \frac{\text{SNR}_{\text{exp},m}^2(\theta)}{\sum_{m=1}^M \text{SNR}_{\text{exp},m}^2(\theta)}, \tag{33}$$

where one can substitute (22) into this equation.

Once the combined image is formed, the overall SNR can be computed via

$$\text{SNR}_{\text{HDR}}(\theta) = \frac{\theta}{\sqrt{\sum_{m=1}^M \left(\frac{w_m}{\tau_m}\right)^2 \sigma_m^2(\theta)}}. \tag{34}$$

where $\sigma_m^2(\theta) = \text{Var}[\mu^{-1}(\overline{Y}^m)]$ is the noise variance of the mth reconstructed LDR image. The intuitive interpretation of (34) is that it weighs the noise according to the integration time τ_m and combination weight w_m to produce a calibrated noise. Thus, the ratio is the SNR with respect to the optimally combined image $\widehat{\theta}$.

The question to be answered here is: If the weight w_m is computed by using the output-referred SNR while the actual sensor has a small full-well capacity such as $L = 7$, what will $\text{SNR}_{\text{HDR}}(\theta)$ be? To answer this question, consider the following configurations: Assume four integration times $\tau_1 = 1$, $\tau_2 = 0.1$, $\tau_3 = 0.01$, $\tau_4 = 0.001$, a total number of frames $N = 100$ at each integration time, and a full-well capacity of $L = 7$. The output-referred SNR for each m, $\text{SNR}_{\text{out},m}$, is computed by (32), whereas the exposure-referred SNR for each m, $\text{SNR}_{\text{exp},m}$, is computed by (22). Once computed, the weight w_m is constructed from $\text{SNR}_{\text{out},m}$. The overall $\text{SNR}_{\text{HDR}}(\theta)$ is formed by using (34) where the calibrated noise $\sigma_m^2(\theta)$ uses the exposure-referred noise $\sigma_{\text{exp},m}^2(\theta)$, essentially $\text{Var}[Y] \cdot \frac{d\mu}{d\theta}$ defined in (22).

Since the weight w_m is computed using SNR_{out} while the actual sensor has a small full-well capacity, the overall SNR_{HDR} will suffer. Figure 9 shows the SNR_{HDR}'s where the weights are either computed from SNR_{out} or SNR_{exp}. This is a new plot that has never been shown including [13]. As one can see, SNR_{HDR} suffers in two places. The first place is the gap between two consecutive maxima, where the weights
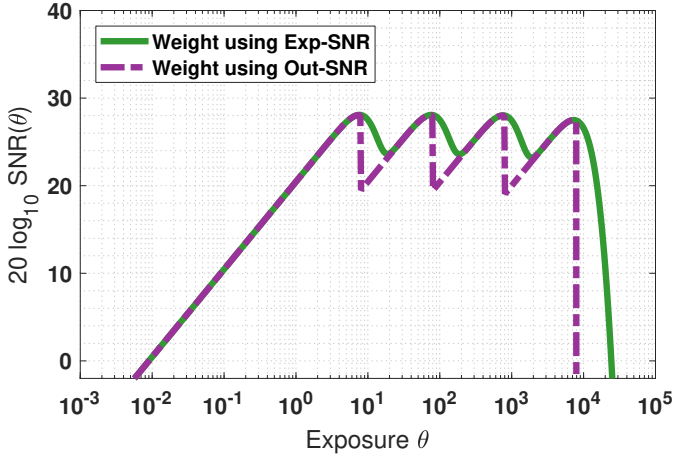
Fig. 9. Comparison of the exposure bracketing by using two different schemes to compute the weight when the sensor has a full-well capacity of $L = 7$. (i) Use the exposure-referred SNR to compute the weight. (ii) Use the output-referred SNR to compute the weight.

generated by the SNR_{out} gives an overall SNR_{HDR} with a sharp cutoff. This discontinuity will be visible when the scene contains a continuous range of θ. The second place is the sharp cutoff after the full-well capacity. The visual impact is that the overall HDR image will saturate sooner than what it is supposed to be, compared to the fusion using exposure-referred SNR.

Figure 10 provides a visual comparison between the HDR image reconstructed using SNR_{out} and SNR_{exp}. In this example, the full-well capacity is assumed to be $L = 7$. Four different exposures (1 ms, 0.1 ms, 0.01 ms, and 0.001 ms) were used to construct the low dynamic range images, where each image is the result of $N = 100$ frames averaged over time. The scene is static, and so the reconstructed results are theoretically optimal with respect to the linear combination and the choice of the SNR. The visual comparison shows a clear benefit of SNR_{exp} over SNR_{out} especially around the cropped areas where the pixels are near saturation. However, the performance gap will become smaller when the full-well capacity becomes larger.

B. Threshold for one-bit Quanta Image Sensors

The second application is the theoretical analysis of the one-bit quanta images sensor (QIS). The one-bit QIS has an interesting forward model given by (6) where the corresponding $\text{SNR}_{\text{exp}}(\theta)$ is derived in (20) (assuming that the read noise σ_{read} is negligible compared to the signal θ.) This section will show two new results on the optimal threshold q that were not mentioned in [5].

1) Optimal q for $\theta \gg 1$: The context of this operating regime is that the sensor sees a sufficient amount of photons but it chooses to operate in a single-bit mode to earn the speed. In this case, the threshold q should be dynamically adjusted to maximize the SNR.

When $\theta \gg 1$, the read noise σ_{read} can be neglected. Consequently, the SNR follows the Poisson statistics as in (20). By noting that $0 \le \Psi_q(\theta) \le 1$ and hence $\Psi_q(\theta)(1-\Psi_q(\theta)) \le 1/4$

(a) Exposures   (b) Ground Truth

(c) Using SNR_{out} (d) Using SNR_{exp}
PSNR = 37.53dB   PSNR = 45.02dB

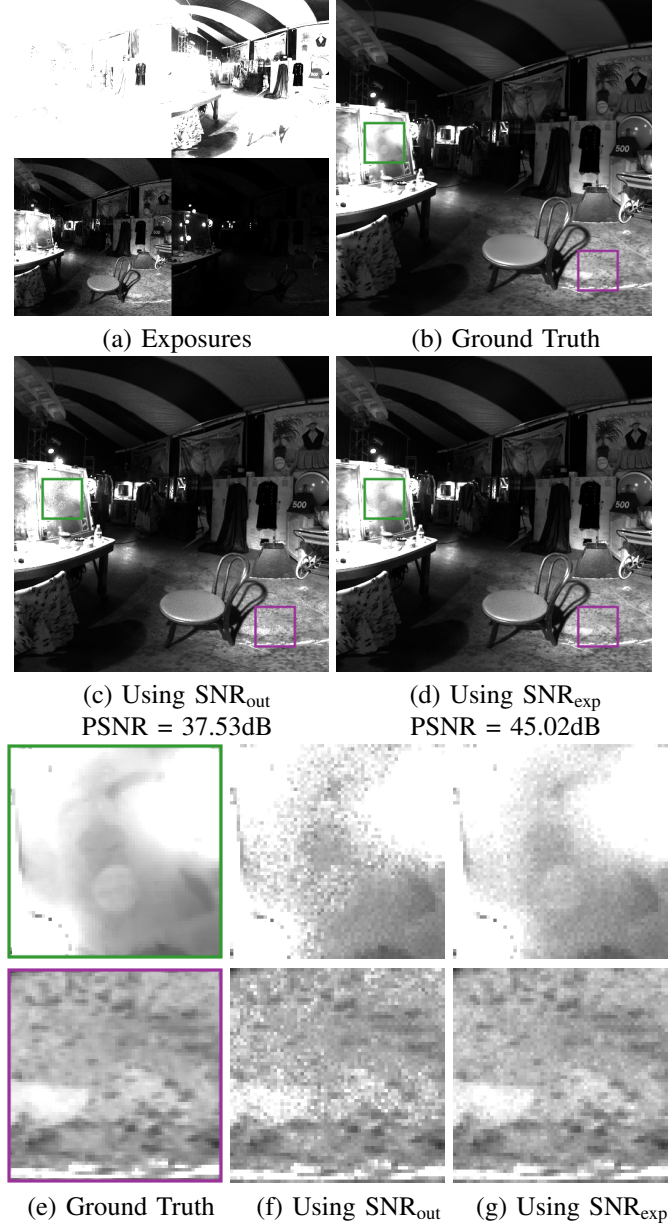(e) Ground Truth   (f) Using SNR_{out} (g) Using SNR_{exp}

Fig. 10. High dynamic range (HDR) image reconstruction using the optimal linear combination scheme. (a) The four exposures are captured at 1 ms, 0.1 ms, 0.01 ms, and 0.001 ms. Shown in this sub-figure are the single-frame measurements with a full-well capacity of 3 bits. (b) The ground truth HDR image. (c) Linear combination using the output-referred SNR. (d) Linear combination using the exposure-referred SNR. For (c)-(d), the reconstruction is based on averaging 100 frames per exposure. (e)-(f)-(g) shows the zoom-in regions of the respective figures.

where the maximum is attained when $\Psi_q(\theta) = 1/2$, $\text{SNR}_{\text{exp}}(\theta)$ can be lower bounded by

$$
\begin{aligned}
\text{SNR}_{\text{exp}}(\theta) &= \frac{\theta}{\sqrt{\Psi_q(\theta)(1-\Psi_q(\theta))}} \cdot \frac{\theta^{q-1}e^{-\theta}}{(q-1)!} \\
&\ge \frac{\theta}{\frac{1}{2}} \cdot \frac{\theta^{q-1}e^{-\theta}}{(q-1)!}.
\end{aligned} \tag{35}
$$

The following lemma shows the optimal q for the lower bound.

Lemma 2 (Maximizing the lower bound): Consider the SNR lower bound for one-bit QIS given by (35). The bound is

maximized when $q^* = \lfloor \theta \rfloor + 1$.

Proof: The proof of this lemma can be found in [5]. $\square$

What was not proved in [5] is that at this optimal q, the inequality in (35) is actually an equality. The reason is that for large θ, $\Psi_q(\theta)$ can be approximated by the cumulative distribution function of a Gaussian via the Stirling's formula. This can be seen from (10) where for large θ,

$$\Psi_q(\theta) \approx \int_{-\infty}^{q} \frac{1}{\sqrt{2\pi\theta}} \exp\left\{ -\frac{(\theta + 1 - y)^2}{2\theta} \right\} dy.$$

The integral is $\Psi_q(\theta) = 1/2$ when $q = \theta + 1$, because the integrand is a Gaussian probability distribution centered at $\theta + 1$. In other words, at the optimal q, $\Psi_q(\theta) = 1/2$ and so $\Psi_q(\theta)(1 - \Psi_q(\theta)) = 1/4$. Hence, the equality in (35) is satisfied, meaning that $q = \theta + 1$ does not only maximize the lower bound but it also maximizes the SNR.

2) Optimal q for $\theta \approx 1$: . This is a new result. The context is that the photon flux is small and so the goal of the sensor is to count the number of photons. However, for small θ, the read noise plays a role because if σ_{read} is big, two adjacent counts cannot be differentiated. The threshold q in this context is used to quantize the analog voltage (which is a Poisson-Gaussian random variable) so that the SNR is maximized.

For one-bit Poisson-Gaussian, the probability distribution of the binary random variable Y still follows (6) but with $X \sim \text{Poisson}(\theta) + \text{Gaussian}(0, \sigma_{\text{read}}^2)$. The exposure-referred SNR is computable because Y is binary. To see this, notice that the mean of Y is

$$\underbrace{\mathbb{E}[Y]}_{\mu(\theta)} = \sum_{k=0}^{\infty} \frac{\theta^k e^{-\theta}}{k!} \int_{-\infty}^{q} \frac{1}{\sqrt{2\pi\sigma_{\text{read}}^2}} \exp\left\{ -\frac{(y-k)^2}{2\sigma_{\text{read}}^2} \right\} dy$$

$$= \frac{1}{2} \sum_{k=0}^{\infty} \frac{\theta^k e^{-\theta}}{k!} \text{erfc}\left(\frac{q-k}{\sqrt{2}\sigma_{\text{read}}} \right), \tag{36}$$

where $\text{erfc}(\cdot)$ is the error function. In the first equation the term inside the summation is the convolution of a Poisson probability density and a Gaussian probability density, evaluated at the threshold q. (36) can be computed using numerical techniques.

The variance of Y follows from the fact that Y is binary (i.e. a Bernoulli random variable). Thus,

$$\text{Var}[Y] = \mathbb{E}[Y](1 - \mathbb{E}[Y]) = \mu(\theta)(1 - \mu(\theta)). \tag{37}$$

The derivative $d\mu/d\theta$ can be approximated numerically:

$$\frac{d\mu}{d\theta} \approx \frac{\mu(\theta + \epsilon) - \mu(\theta)}{\epsilon}, \tag{38}$$

where ϵ is a small numerical constant. Combining everything in (36), (37) and (38), the exposure-referred SNR can be numerically computed via Theorem 1.

The derived $\text{SNR}_{\text{exp}}(\theta)$ is plotted in Figure 11 as a function of the threshold q, for various read noise levels σ_{read}. When $\sigma_{\text{read}} = 0$, the SNR is a constant for all q considered. This is because in the absence of read noise, there is no ambiguity in determining the measured number of photons regardless where the threshold q is put. As σ_{read} increases, two adjacent counts begin to overlap due to the Gaussian. The maximum is located around $q = 0.5$. For high read noise where the two peaks of
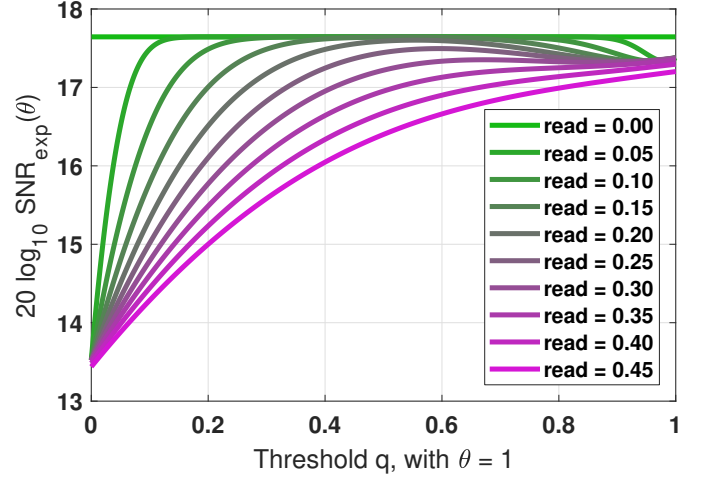

Fig. 11. SNR_{exp} as a function of the threshold when the exposure is $\theta = 1$. As the read noise σ_{read}, increases, the SNR drops when $q = 0.5$.

the photon count merge, there is limited SNR one can expect from the sensor.

VII. DISCUSSIONS AND CONCLUSION

A. Alternatives to SNR?

While SNR is a natural choice for analyzing the performance of an image sensor, it is by no means the *only* option. Especially for one-bit devices such as the quanta image sensor, there are other ways to characterize the performance.

1) Entropy: As far as one-bit measurements are concerned, the entropy is a natural substitute of the SNR. If Y is binary with $p_Y(1) = 1 - \Psi_q(\theta)$ and $p_Y(0) = \Psi_q(\theta)$, the entropy is

$$H(Y) = -p_Y(1) \log_2 p_Y(1) - p_Y(0) \log_2 p_Y(0)$$
$$= -(1 - \Psi_q(\theta)) \log_2(1 - \Psi_q(\theta)) - \Psi_q(\theta) \log_2 \Psi_q(\theta).$$

It is relatively easy to show that the derivative of the entropy with respect to $\Psi_q(\theta)$ is

$$\frac{d}{d\Psi_q(\theta)} H(Y) = -\log\left(\frac{1 - \Psi_q(\theta)}{\Psi_q(\theta)} \right).$$

Setting it to zero will yield $\Psi_q(\theta) = \frac{1}{2}$. Therefore, the entropy is maximized when $\mathbb{E}[Y] = 1 - \Psi_q(\theta) = \frac{1}{2}$. Since $\mathbb{E}[Y]$ is the expected value of the measurement, $\mathbb{E}[Y] = \frac{1}{2}$ means that the entropy is maximized when there are 50% one's and 50% zero's in a set of independent measurements. So, if the application goal is to identify a threshold q such that the performance of the sensor is maximized, then instead of optimizing for the SNR as in Section VI.B.1, the alternative is to optimize the entropy. The solution to $\Psi_q(\theta) = 1/2$ is $q^* = \theta + 1$, which is consistent with Section VI.B.1.

2) Bit Error Rate (BER): In the presence of read noise, the bit error rate is another commonly used criterion to evaluate the performance of a sensor. For one-bit quanta image sensor, the BER measures the probability of making a wrong decision

(i.e., declaring a 0 as a 1, or declaring a 1 as a 0). It can be readily computed as

$$
\begin{aligned}
\text{BER}(\theta) &= p_Y(0) \cdot \int_q^\infty \frac{1}{\sqrt{2\pi\sigma_{\text{read}}^2}} e^{-\frac{t^2}{2\sigma_{\text{read}}^2}}\, dt \\
&\quad + p_Y(1) \cdot \int_{-\infty}^q \frac{1}{\sqrt{2\pi\sigma_{\text{read}}^2}} e^{-\frac{(t-1)^2}{2\sigma_{\text{read}}^2}}\, dt \\
&= \frac{1}{2}\text{erfc}\left(\frac{q}{\sigma_{\text{read}}\sqrt{2}}\right)\Psi_q(\theta) \\
&\quad + \frac{1}{2}\text{erfc}\left(\frac{1-q}{\sigma_{\text{read}}\sqrt{2}}\right)(1-\Psi_q(\theta)). \qquad (39)
\end{aligned}
$$

Therefore, if $q = 1/2$, the BER is simplified to

$$
\text{BER}(\theta) = \frac{1}{2}\text{erfc}\left(\frac{1}{\sigma_{\text{read}}\sqrt{8}}\right), \qquad (40)
$$

which does not depend on θ. If $\text{BER}(\theta)$ can be empirically measured, then by inverting (40) one can estimate the read noise σ_{read}. For a fixed θ, one can also optimize (39) by finding an appropriate q.

B. Conclusion

The exposure-referred SNR is a concept motivated by the need to capture the sensor's behavior near and beyond the full-well capacity. For small image sensors with one or few bits, exposure-referred SNR provides a natural characterization of the performance without showing an infinite SNR due to the artificial squeezing of the noise. In order to establish the exposure-referred SNR, the paper introduces new mathematical concepts and showed a few results:

- The mean invariance property is introduced. The property asserts that when an estimator $\widehat{\theta}(\cdot)$ is applied to the mean of the measurement, the mapped value is the true parameter $\widehat{\theta}(\mu) = \theta$.
- Exposure-referred SNR calibrates the noise variance $\text{Var}[Y]$ by the derivative $d\mu/d\theta$, where the derivative is the result of the first-order approximation used in the Delta Method.
- SNR of a sensor with a finite full-well capacity is analytically derived via the incomplete Gamma function. The new result generalizes the conventional ones. For sensors with a large full-well capacity, the new result recovers the classical one that shows a linear response. For sensors with a small full-well capacity, the new result shows how the transient of the SNR looks like.
- Monte Carlo simulation is presented for complex noise models. The general procedure is to estimate the mean of the measurement $\mu(\theta)$ and the variance of the measurement $\text{Var}[Y]$. Then by using a numerical finite difference operator, the derivative $d\mu/d\theta$ can be approximated.
- Optimal high dynamic range image reconstruction is shown. The new result generalizes the classical linear response.
- Threshold analysis of one-bit sensors is revisited and generalized. For $\theta \gg 1$, the optimal threshold is $q = \theta+1$. For $\theta \approx 1$, the optimal threshold needs to overcome the read noise and so the optimal value is $q = 1/2$.

- For one-bit sensors, it was found that the entropy of the bits and the bit error rate can be used as alternatives to characterize the sensor.

As the full-well capacity of the image sensors is becoming small, it is anticipated that the exposure-referred SNR will become a useful utility for theoretical analysis of the sensors, hence allowing signal processing [23]–[26], algorithm development [27]–[31], and computer vision applications [32], [33]. Specific problems such as sensor gain control, exposure analysis, bit-depth to speed trade off, and color filter array, will be important questions to answer next.

VIII. APPENDIX

A. Gaussian approximation of Poisson

When deriving the first-order derivative of the incomplete Gamma function, it was mentioned that the Poisson distribution can be approximated by a Gaussian. The formal statement is as follows.

Lemma 3 (Gaussian approximation of Poisson): For large θ (i.e., $\theta \gg 1$), it holds that

$$
p_X(x) \stackrel{\text{def}}{=} \frac{\theta^x e^{-\theta}}{x!} \approx \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\theta)^2}{2\theta}}. \qquad (41)
$$

Note that this is *not* the Central Limit theorem because it does not involve any sample average. The approximation compares the two functions.

Proof: First of all, take the log on the Poisson equation:

$$
\log p_X(x) = \log\left\{\frac{\theta^x e^{-\theta}}{x!}\right\} = x\log\theta - \theta - \log x!
$$

Stirling's formula states that for $x \to \infty$, we have $x! \approx x^x e^{-x}\sqrt{2\pi x}$. Substitute into the previous equation yields

$$
\begin{aligned}
\log p_X(x) &\approx x\log\theta - \theta - \log\left(x^x e^{-x}\sqrt{2\pi x}\right) \\
&= x\log\theta - \theta - x\log x + x - \log\sqrt{2\pi x}.
\end{aligned}
$$

The Gaussian has to fit the Poisson well around the mean, which is θ. Thus define $x = \theta + \epsilon$ with $\theta \gg \epsilon$. Then,

$$
\begin{aligned}
\log p_X(x) &= x\log\theta - \theta - x\log x + x - \log\sqrt{2\pi x} \\
&= (\theta+\epsilon)\log\theta - \theta - (\theta+\epsilon)\log(\theta+\epsilon) \\
&\quad + (\theta+\epsilon) - \log\sqrt{2\pi(\theta+\epsilon)} \\
&= \epsilon + (\theta+\epsilon)\log\frac{\theta}{\theta+\epsilon} - \log\sqrt{2\pi(\theta+\epsilon)} \\
&= \epsilon - (\theta+\epsilon)\log\left(1+\frac{\epsilon}{\theta}\right) - \log\sqrt{2\pi\theta} - \frac{1}{2}\log\left(1+\frac{\epsilon}{\theta}\right) \\
&= \epsilon - \log\sqrt{2\pi\theta} - \left(\theta+\epsilon+\frac{1}{2}\right)\log\left(1+\frac{\epsilon}{\theta}\right).
\end{aligned}
$$

For $\frac{\epsilon}{\theta} \ll 1$, it holds that $\log(1+\frac{\epsilon}{\theta}) \approx \frac{\epsilon}{\theta} - \frac{\epsilon^2}{2\theta^2} + \ldots$. Therefore,

$$\log p_X(x) \approx \epsilon - \log\sqrt{2\pi\theta} - \left(\theta + \epsilon + \frac{1}{2}\right)\left(\frac{\epsilon}{\theta} - \frac{\epsilon^2}{2\theta^2} + \ldots\right)$$

$$= \epsilon - \log\sqrt{2\pi\theta} - \epsilon - \frac{\epsilon^2}{\theta} - \frac{\epsilon}{2\theta} + \frac{\epsilon^2}{2\theta} + \frac{\epsilon^2}{4\theta^2} + \ldots.$$

By canceling terms, and removing $\frac{\epsilon^2}{4\theta^2}$ and $\frac{\epsilon}{2\theta}$ (because $\frac{\epsilon}{\theta} \ll 1$), it follows that $\log p_X(x) \approx -\frac{\epsilon^2}{2\theta} - \log\sqrt{2\pi\theta}$. This implies that $p_X(x) \approx \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{\epsilon^2}{2\theta}}$. Substituting $x = \theta + \epsilon$ completes the proof. $\qquad\square$

B. Mean Invariant = ML for Exponential Family

A mean invariant estimator is generally not the same as an ML estimator. However, for distributions in the exponential family, they are identical. Recall the definition of the exponential family. A sequence of i.i.d. random variables $\boldsymbol{Y} = [Y_1, \ldots, Y_N]$ is said to be in the exponential family if

$$p_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = h(\boldsymbol{y})\exp\left\{\eta(\theta)T(\boldsymbol{y}) - A(\theta)\right\}, \qquad (42)$$

for some functions h, η, T, and A [12]. The function $T(\boldsymbol{y})$ is the sufficient statistic. Since a necessary condition for an estimator to be mean invariant is that it is a function of the sample average $\overline{Y}$, in the subsequent discussions it is assumed that $T(\boldsymbol{y}) = \overline{y}$ where $\overline{y} = (1/N)\sum_{n=1}^{N} y_n$ with $y_1, \ldots, y_N$ as the realizations of $Y_1, \ldots, Y_N$.

Example 9: For one-bit QIS, the probability density function can be written as

$$p_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \prod_{n=1}^{N}\exp\left\{\log\left[(1 - e^{-\theta})^{y_n}(e^{-\theta})^{1-y_n}\right]\right\}$$

$$= \exp\left\{N\overline{y}\log(1 - e^{-\theta}) + N(1 - \overline{y})\log(e^{-\theta})\right\}$$

$$= \left[\exp\left\{\overline{y}\log(e^{\theta} - 1) - \theta\right\}\right]^{N}.$$

So, one can associate $h(\boldsymbol{y}) = 1$, $T(\boldsymbol{y}) = \overline{y}$, $\eta(\theta) = \log(e^{\theta}-1)$, $A(\theta) = \theta$.

Lemma 4: $\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{y})$ is the ML estimate if and only if it satisfies the equation

$$g(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{y})) \overset{\mathrm{def}}{=} \frac{A'(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{y}))}{\eta'(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{y}))} = \overline{y}, \qquad (43)$$

where $g(\theta) \overset{\mathrm{def}}{=} A'(\theta)/\eta'(\theta)$ is a function, provided that $A'(\theta)$ and $\eta'(\theta)$ exist and are not zero. If g^{-1} exists, then $\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{y}) = g^{-1}(\overline{y})$.

Proof: $\widehat{\theta}_{\mathrm{ML}}$ is the ML estimate if and only if $\frac{\partial}{\partial\theta}\log p_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)\big|_{\theta=\widehat{\theta}_{\mathrm{ML}}} = 0$. The derivative is

$$\frac{\partial}{\partial\theta}\log p_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = \frac{\partial}{\partial\theta}\left\{\log h(\boldsymbol{y}) + \eta(\theta)\overline{y} - A(\theta)\right\}$$

$$= \eta'(\theta)\overline{y} - A'(\theta).$$

Setting it to zero, the condition is equivalent to $\frac{A'(\widehat{\theta}_{\mathrm{ML}})}{\eta'(\widehat{\theta}_{\mathrm{ML}})} = \overline{y}$. Defining $g(\theta) = \frac{A'(\theta)}{\eta'(\theta)}$, it follows that $g(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{y})) = \overline{y}$. If g^{-1} exists, then estimator is $\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{Y}) = g^{-1}(\overline{Y})$. $\qquad\square$

Lemma 5: For any distributions in the exponential family and for any parameter θ,

$$\frac{A'(\theta)}{\eta'(\theta)} = \mathbb{E}[T(\boldsymbol{Y})] \overset{\mathrm{def}}{=} \mu(\theta). \qquad (44)$$

If $A'(\theta)$ and $\eta'(\theta)$ exist and are not zero, and μ^{-1} exists, then $\theta = \mu^{-1}(\mathbb{E}[T(\boldsymbol{Y})])$.

Proof: The function $A(\theta)$ is the normalization constant, defined as $A(\theta) = \log\int h(\boldsymbol{y})e^{\eta(\theta)T(\boldsymbol{y})}d\boldsymbol{y}$. Therefore, its derivative is

$$A'(\theta) = \frac{d}{d\theta}\left\{\log\int h(\boldsymbol{y})e^{\eta(\theta)T(\boldsymbol{y})}d\boldsymbol{y}\right\}$$

$$= \frac{\int h(\boldsymbol{y})e^{\eta(\theta)T(\boldsymbol{y})}T(\boldsymbol{y})\eta'(\theta)d\boldsymbol{y}}{\int h(\boldsymbol{y})e^{\eta(\theta)T(\boldsymbol{y})}\boldsymbol{y}}$$

$$= \eta'(\theta)\int T(\boldsymbol{y})\cdot p_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)d\boldsymbol{y} = \eta'(\theta)\mathbb{E}[T(\boldsymbol{Y})].$$

Thus, if μ^{-1} exists, then $\theta = \mu^{-1}(\mathbb{E}[T(\boldsymbol{Y})])$. $\qquad\square$

Comparing (43) and (44), it follows that the mapping is $g(\theta) = \mu(\theta)$. Therefore, $\mu(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{Y})) = g(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{Y}))$. But since $g(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{Y})) = \overline{Y}$, it follows that $\mu(\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{Y})) = \overline{Y}$. Thus, the ML estimator is mean invariant.

C. Proof of Theorem 2

Proof: Since $\overline{Y} = (1/N)\sum_{n=1}^{N} Y_n$, the mean $\mathbb{E}[\overline{Y}] = \mathbb{E}[Y_1]$. It then follows that the derivative $d\mu/d\theta$ remains unchanged. For the variance, it is easy to show that $\mathrm{Var}[\overline{Y}] = \mathrm{Var}[Y_1]/N$. Substituting these results into (22) would yield

$$\mathrm{SNR}_{\exp}(\theta) = \frac{\theta}{\sqrt{\mathrm{Var}[\overline{Y}]}}\cdot\frac{d\mu}{d\theta} = \sqrt{N}\frac{\theta}{\sqrt{\mathrm{Var}[Y_1]}}\cdot\frac{d\mu}{d\theta}.$$

Therefore, it suffices to prove the SNR for $N = 1$. In what remains, let Y be any of the random variables $Y_1, \ldots, Y_N$ since they are i.i.d.

Recall the probability density function of Y:

$$p_Y(y) = \begin{cases} \frac{\theta^y}{y!}e^{-\theta}, & y < L, \\ \sum_{k=L}^{\infty}\frac{\theta^k}{k!}e^{-\theta} = 1 - \Psi_L(\theta), & y \geq L, \end{cases}$$

where $\Psi_L(\theta)$ is the incomplete Gamma function. The mean of Y can be shown as

$$\mu = \mathbb{E}[Y] = \sum_{k=0}^{L-1} k\cdot\frac{\theta^k}{k!}e^{-\theta} + L\cdot\left(\sum_{k=L}^{\infty}\frac{\theta^k}{k!}e^{-\theta}\right)$$

$$= \sum_{k=1}^{L-1}\frac{\theta^k}{(k-1)!}e^{-\theta} + L\cdot(1 - \Psi_L(\theta))$$

$$= \theta\sum_{k=0}^{L-2}\frac{\theta^{k-1}}{k!}e^{-\theta} + L\cdot(1 - \Psi_L(\theta))$$

$$= \theta\Psi_{L-1}(\theta) + L\cdot(1 - \Psi_L(\theta)).$$

The derivative $d\mu/d\theta$ is therefore

$$\frac{d\mu}{d\theta} = \frac{d}{d\theta}\left\{\theta\Psi_{L-1}(\theta) + L\cdot(1 - \Psi_L(\theta))\right\}$$

$$= \theta\Psi'_{L-1}(\theta) + \Psi_{L-1}(\theta) - L\cdot\Psi'_L(\theta).$$

For the variance, since $\text{Var}[Y] = \mathbb{E}[Y^2] - \mu^2$, it remains to determine $\mathbb{E}[Y^2]$.

$$
\begin{aligned}
\mathbb{E}[Y^2] &= \sum_{k=0}^{L-1} k^2 \cdot \frac{\theta^k}{k!} e^{-\theta} + L^2 \cdot \left(\sum_{k=L}^{\infty} \frac{\theta^k}{k!} e^{-\theta} \right) \\
&= \sum_{k=1}^{L-1} k \frac{\theta^k}{(k-1)!} e^{-\theta} + L^2 \cdot (1 - \Psi_L(\theta)) \\
&= \sum_{k=1}^{L-1} (k-1+1) \frac{\theta^k}{(k-1)!} e^{-\theta} + L^2 \cdot (1 - \Psi_L(\theta)) \\
&= \sum_{k=2}^{L-1} \frac{\theta^k}{(k-2)!} e^{-\theta} + \sum_{k=1}^{L-1} \frac{\theta^k}{(k-1)!} e^{-\theta} \\
&\qquad + L^2 \cdot (1 - \Psi_L(\theta)) \\
&= \theta^2 \Psi_{L-2}(\theta) + \theta \Psi_{L-1}(\theta) + L^2 (1 - \Psi_L(\theta)).
\end{aligned}
$$

This completes the proof. $\qquad \square$

D. Proof of Corollary 2

Proof: When L is large, $\Psi_L(10^\phi)$ and $\Psi_{L-1}(10^\phi)$ are close enough that they can be considered approximately equal. Denote the value $\Psi_L(10^\phi)$ as Ψ. Then by (24) it holds that $\Psi \to 0$ for $\phi > \log_{10} L$ and $\Psi \to 1$ for $\phi \le \log_{10} L$. In either case, since Ψ is a constant, it follows that the derivative $\Psi'_L(10^\phi) = 0$ as long as $\phi > \log_{10} L$ or $\phi < \log_{10} L$. Therefore, by denoting $\theta = 10^\phi$, the two cases can be derived as follows.

When $\phi \le \log_{10} L$, it holds that

$$
\begin{aligned}
\mu &= \theta \Psi + L(1 - \Psi) = \theta, \\
\text{Var}[Y] &= \theta^2 \Psi + \theta \Psi + L^2(1 - \Psi) - \mu^2 \\
&= \theta^2 \cdot 1 + \theta \cdot 1 + L^2 \cdot 0 - \mu^2 = \theta, \\
\frac{d\mu}{d\theta} &= \Psi - (L - \theta)\Psi'_{L-1}(\theta) \\
&= 1 - (L - \theta) \cdot 0 = 1.
\end{aligned}
$$

So, the SNR for $\phi \le \log_{10} L$ is

$$
20 \log_{10} \text{SNR}_{\exp}(10^\phi) = 20 \log_{10} \sqrt{10^\phi} = 10\phi.
$$

When $\phi > \log_{10} L$, $\Psi \to 0$. Therefore,

$$
\begin{aligned}
\mu &= \theta \Psi + L(1 - \Psi) \approx \theta \Psi + L, \\
\text{Var}[Y] &= \theta^2 \Psi + \theta \Psi + L^2(1 - \Psi) - \mu^2 \\
&= \theta^2 \Psi + \theta \Psi + L^2 - (\theta\Psi + L)^2 \\
&= \theta^2 \Psi + \theta \Psi + L^2 - \theta^2 \Psi^2 - 2\theta\Psi L - L^2 \\
&= \theta^2 \Psi + \theta \Psi - 2\theta\Psi L \\
&= \theta\Psi(\theta + 1 - 2L), \\
\frac{d\mu}{d\theta} &= \theta\Psi'_{L-1}(\theta) + \Psi_{L-1}(\theta) - L\Psi'_L(\theta) \\
&= \theta \cdot 0 + \Psi - L \cdot 0 = \Psi.
\end{aligned}
$$

By taking the limit that $\Psi \to 0$, it follow that

$$
\begin{aligned}
\lim_{\Psi \to 0} \text{SNR}_{\exp}(10^\phi) &= \lim_{\Psi \to 0} \frac{\theta}{\sqrt{\text{Var}[Y]}} \cdot \frac{d\mu}{d\theta} \\
&= \lim_{\Psi \to 0} \frac{\theta}{\sqrt{\theta\Psi(\theta + 1 - 2L)}} \cdot \Psi \\
&= \lim_{\Psi \to 0} \frac{\sqrt{\theta\Psi}}{\sqrt{(\theta + 1 - 2L)}} = 0.
\end{aligned}
$$

Combining with the case where $\phi \le \log_{10} L$, the overall SNR is proved. $\qquad \square$

E. MATLAB Code for Monte Carlo Simulation

The MATLAB code below illustrates the Monte Carlo simulation of how $\text{SNR}_{\exp}(\theta)$ is generated for a truncated Poisson distribution. Adding other factors to the forward model can be done by modifying the random variable Y.

```
N     = 100000;
L     = 10;
theta_set = logspace(-2,3,100);
mu    = zeros(1,100);
sigma = zeros(1,100);
for i=1:100
    theta    = theta_set(i);
    Theta    = theta*ones(N,1);
    Y        = poissrnd(Theta);
    Y(Y>L)   = L;
    mu(i)    = mean(Y);
    sigma(i) = std(Y);
end
dmu_dt  = [diff(mu)./diff(theta_set) 1];
SNR     = theta_set./sigma.*dmu_dt;
loglog(theta_set, SNR);
```

For plotting the theoretical $\text{SNR}_{\exp}(\theta)$, one just needs to call the incomplete Gamma function.

```
theta = logspace(-2,3,100);
Psi   = gammainc(theta,L,'upper');
Psi1  = gammainc(theta,L-1,'upper');
Psi2  = gammainc(theta,L-2,'upper');
dPsi  = -theta.^(L-1).*exp(-theta) ...
        /gamma(L);
dPsi1 = -theta.^(L-2).*exp(-theta) ...
        /gamma(L-1);
mu    = theta.*Psi1 + L.*(1 - Psi);
sigma = sqrt(theta.^2 .* Psi2 + ...
        theta.*Psi1 + L^2*(1-Psi) - ...
        mu_theory.^2);
dmu_dt = theta.*dPsi1 + Psi1 - L*dPsi;
SNR    = theta./sigma.*dmu_dt;
loglog(theta, SNR);
```

The combination of the two pieces of codes, with minor modifications, is sufficient to reproduce all the figures reported in this paper.

References

[1] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[2] E. M. V. Association, "EMVA Standard 1288: Standard for characterization of image sensors and cameras," 2010. Available online at https://www.emva.org/wp-content/uploads/EMVA1288-3.0.pdf.

[3] A. E. Gamal, "EE392B Lecture Note: SNR and Dynamic Range." https://isl.stanford.edu/~abbas/ee392b/lect08.pdf, accessed 12/7/2021.

[4] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *IEEE CVPR*, vol. 1, pp. 374–380, 1999.

[5] O. A. Elgendy and S. H. Chan, "Optimal threshold design for Quanta Image Sensor," *IEEE Trans. Computational Imaging*, vol. 4, no. 1, pp. 99–111, 2018.

[6] E. R. Fossum, "Modeling the performance of single-bit and multi-bit quanta image sensors," *IEEE J. Electron Devices Society*, vol. 1, no. 9, pp. 166–174, 2013.

[7] F. Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli, "Bits from photons: Oversampled image acquisition using binary Poisson statistics," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1421–1436, 2012.

[8] J. Nakamura, *Image Sensors and Signal Processing for Digital Still Cameras*. CRC Press, Talyor and Francis Group, 2005.

[9] S.-H. Lim, "Characterization of noise in digital photographs for image processing," in *Digital Photography II*, vol. 6069, p. 60690O, International Society for Optics and Photonics, 2006.

[10] E. R. Fossum, "Some thoughts on future digital still cameras," in *Image Sensors and Signal Processing for Digital Still Cameras*, p. 305, CRC, 2006.

[11] J. R. B. Whittlesey, "Incomplete gamma functions for evaluating erlang process probabilities," *Mathematics of Computation*, vol. 17, no. 81, pp. 11–17, 1963.

[12] E. L. Lehmann, *Elements of Large-Sample Theory*. Springer, 1999.

[13] A. Gnanasambandam and S. H. Chan, "HDR imaging with Quanta Image Sensors: Theoretical limits and optimal reconstruction," *IEEE Trans. Computational Imaging*, vol. 6, pp. 1571–1585, 2020.

[14] A. Ingle, A. Velten, and M. Gupta, "High flux passive imaging with single-photon sensors," in *IEEE CVPR*, pp. 6760–6769, 2019.

[15] P. McCullagh and J. A. Nelder, *Generalized linear models*. Chapman and Hall, 1983.

[16] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. Lensch, "Optimal HDR reconstruction with linear digital cameras," in *IEEE CVPR*, pp. 215–222, 2010.

[17] A. El Gamal and H. Eltoukhy, "CMOS image sensors," *IEEE Circuits and Devices Magazine*, vol. 21, no. 3, pp. 6–20, 2005.

[18] J. Ma, S. Masoodian, D. A. Starkey, and E. R. Fossum, "Photon-number-resolving megapixel image sensor at room temperature without avalanche gain," *OSA Optica*, vol. 4, pp. 1474–1481, Dec 2017.

[19] S. Mann and R. Picard, "Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," Tech. Rep. 323, M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts, 1994.

[20] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH*, pp. 369–378, 1997.

[21] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *IEEE CVPR*, vol. 1, pp. 472–479, 2000.

[22] K. Kirk and H. J. Andersen, "Noise characterization of weighting schemes for combination of multiple exposures," in *British Machine Vision Conf.*, pp. 1129–1138, 2006.

[23] A. Gnanasambandam, O. Elgendy, J. Ma, and S. H. Chan, "Megapixel photon-counting color imaging using Quanta Image Sensor," *OSA Optics Express*, vol. 27, no. 12, pp. 17298–17310, 2019.

[24] O. A. Elgendy and S. H. Chan, "Color Filter Arrays for Quanta Image Sensors," *IEEE Trans. Computational Imaging*, vol. 6, pp. 652–665, Jan 2020.

[25] O. A. Elgendy, A. Gnanasambandam, S. H. Chan, and J. Ma, "Low-light demosaicking and denoising for small pixels using learned frequency selection," *IEEE Transa. Computational Imaging*, vol. 7, pp. 137–150, 2021.

[26] S. H. Chan, "On the insensitivity of bit density to read noise in one-bit Quanta Image Sensors," 2022. Available online at http://arxiv.org/abs/2203.06086, accessed 6/12/2022.

[27] S. H. Chan and Y. M. Lu, "Efficient image reconstruction for gigapixel Quantum Image Sensors," in *IEEE Global Conf. Signal and Info. Process.*, pp. 312–316, 2014.

[28] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Computational Imaging*, vol. 3, pp. 84–98, Nov. 2016.

[29] S. H. Chan, O. A. Elgendy, and X. Wang, "Images from bits: Non-iterative image reconstruction for Quanta Image Sensors," *MDPI Sensors*, vol. 16, no. 11, p. 1961, 2016.

[30] J. H. Choi, O. A. Elgendy, and S. H. Chan, "Image reconstruction for Quanta Image Sensors using deep neural networks," in *IEEE ICASSP*, pp. 6543–6547, 2018.

[31] Y. Chi, A. Gnanasambandam, V. Koltun, and S. H. Chan, "Dynamic low-light imaging with Quanta Image Sensors," in *ECCV*, pp. 122–138, 2020.

[32] A. Gnanasambandam and S. H. Chan, "Image classification in the dark using Quanta Image Sensors," in *ECCV*, pp. 484–501, 2020.

[33] C. Li, X. Qu, A. Gnanasambandam, O. A. Elgendy, J. Ma, and S. H. Chan, "Photon-limited object detection using non-local feature matching and knowledge distillation," in *ICCV-W*, pp. 3959–3970, 2021.