PCH-EM: A Solution to Information Loss in the Photon Transfer Method

Aaron J. Hendrickson, David P. Haefner, Stanley H. Chan, *Senior Member, IEEE*, Nicholas R. Shade, *Graduate Student Member, IEEE*, and Eric R. Fossum, *Life Fellow, IEEE*

***Abstract*— Working from a Poisson–Gaussian noise model, a multisample extension of the photon counting histogram expectation–maximization (PCH-EM) algorithm is derived as a general-purpose alternative to the photon transfer (PT) method. This algorithm is derived from the same model, requires the same experimental data, and estimates the same sensor performance parameters as the time-tested PT method, all while obtaining lower uncertainty estimates. It is shown that as read noise becomes large, multiple data samples are necessary to capture enough information about the parameters of a device under test, justifying the need for a multisample extension. An estimation procedure is devised consisting of initial PT characterization followed by repeated iteration of PCH-EM to demonstrate the improvement in estimating uncertainty achievable with PCH-EM, particularly in the regime of deep subelectron read noise (DSERN). A statistical argument based on the information theoretic concept of sufficiency is formulated to explain how PT data reduction procedures discard information contained in raw sensor data, thus explaining why the proposed algorithm is able to obtain lower uncertainty estimates of key sensor performance parameters, such as read noise and conversion gain. Experimental data captured from a CMOS quanta image sensor with DSERN are then used to demonstrate the algorithm's usage and validate the underlying theory and statistical model. In support of the reproducible research effort, the code associated with this work can be obtained on the MathWorks file exchange (FEX) (Hendrickson et al., 2024).**

***Index Terms*— Conversion gain, deep subelectron read noise (DSERN), expectation–maximization (EM) algorithm, PCH, photon counting, photon counting histogram EM (PCH-EM), photon transfer (PT), QIS, quanta exposure, read noise.**

Manuscript received 16 May 2024; accepted 10 June 2024. Date of publication 24 June 2024; date of current version 25 July 2024. The work of Aaron J. Hendrickson was supported by the Naval Innovative Science and Engineering (NISE) under Project 219BAR-24-043. The work of Nicholas R. Shade was supported by the National Science Foundation under Grant 2125733. The review of this article was arranged by Editor L. Pancheri. *(Corresponding author: Aaron J. Hendrickson.)*

Aaron J. Hendrickson is with U.S. Navy (NAWCAD), Patuxent River, MD 20670 USA (e-mail: aaron.j.hendrickson2.civ@us.navy.mil).

David P. Haefner is with U.S. Army Combat Capabilities Development Command (DEVCOM) C5ISR Center, Fort Belvoir, VA 22060 USA.

Stanley H. Chan is with the School of Electrical and Computer Engineering, Purdue University, Lafayette, IN 47907 USA (e-mail: stanchan@purdue.edu).

Nicholas R. Shade and Eric R. Fossum are with the Thayer School of Engineering, Dartmouth College, Hanover, NH 03755 USA (e-mail: nicholas.r.shade.th@dartmouth.edu; eric.r.fossum@dartmouth.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2024.3414369.

Digital Object Identifier 10.1109/TED.2024.3414369

I. Introduction

EVER since the apparent identification of photon shot noise in early vidicon tubes employed in NASA's space missions, researchers have sought enhanced and more precise techniques for characterizing the continually advancing landscape of imaging technologies [1], [2], [3]. In 1969, Smith and Boyle's CCD image sensor emerged as a groundbreaking invention; a technology later made practical by Tompsett et al.'s frame transfer implementation in 1971 [4], [5], [6]. These devices exhibited notably diminished read noise compared with vidicon tubes, facilitating the distinct observation of shot noise [2]. During this period, Janesick and Elliott popularized the photon transfer (PT) methodology for the characterization of CCDs, a technique later applied to CMOS image sensor technology and integrated into standards, such as the EMVA 1288 [7].

At the core of the PT methodology lies the fundamental principle that raw sensor data captured over many exposure levels should be distilled into means (signal measurements) and variances (noise-squared measurements), wherein the relationships between these sample moments carry information about important sensor performance parameters, such as read noise and conversion gain. From a historical viewpoint, the development of PT coincided with an era where read noise was on the order of multiple electrons. At this level of read noise, the histograms produced by a pixel provided with constant illumination are effectively normally distributed. Given that the sample mean and variance serve as the complete minimal sufficient statistic of the normal model, this PT data reduction strategy can be viewed, from a post hoc perspective, as the optimal approach at high noise, reducing the raw data to a set of summary statistics, all while preserving the information about the unknown sensor parameters contained in it.

In 2004, Fossum conceived the quanta image sensor as a paradigm shift in digital imaging [8], [9]. In 2015, the first CMOS QIS was realized, demonstrating the ability to perform accurate photon number resolving by attaining deep subelectron read noise (DSERN) [10]. This DSERN was accomplished through a combination of reducing the capacitance of the floating diffusion sense node, thereby increasing V/e− gain, and reducing source follower noise [10], [11], [12], [13], [14]. Consequently, CMOS QISs married accurate photon number resolving, facilitated by DSERN, with the high-speed readout and radiation hardness of CMOS technology. Since

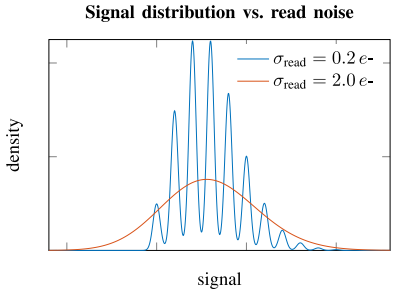

Fig. 1. Pixel signal distribution for small and large read noise.

the advent of CMOS QIS, DSERN has been achieved by various means in commercial products from BAE Systems, Hamamatsu photontics, and Gigajot Technology Inc., while also being experimentally realized in CCD and CMOS architectures via nondestructive "skipper" readout modes [15], [16], [17], [18].

As a result of DSERN, the histograms generated by QIS pixels exhibit markedly nonnormal distributions, featuring structures previously hidden by the relatively large read noise present in traditional CCD and CMOS devices. For example, Fig. 1 shows the signal distribution for the same pixel at two different levels of read noise (σ_{read}). At large (multielectron) noise, the distribution is effectively normal, whereas at small (deep subelectron) noise, periodic structures appear in the distribution. The presence of these periodic structures corresponds to higher information content contained in the sensor data, which cannot be fully captured by sample means and variances. Subsequently, following 2015, new characterization methods were discovered that could leverage this additional information to construct lower uncertainty estimators of the relevant sensor parameters compared with what is possible with the PT methodology [16], [19], [20], [21], [22], [23].

Among these nascent techniques for characterization, Hendrickson and Haefner's photon counting histogram expectation–maximization (PCH-EM) algorithm displayed potential, enabling the computation of maximum likelihood estimates (MLEs) of the sensor parameters from a single data sample when the sensor being tested operates within the DSERN regime [22], [24]. Drawing inspiration from the PT method, PCH-EM can be generalized to integrate data from multiple samples taken at distinct exposure levels, enhancing the precision of the parameter estimates and extending the algorithm outside the DSERN regime. Remarkably, this multisample extension of PCH-EM utilizes the same experiment and estimates the same parameters as PT, all while delivering estimates with reduced uncertainty and superior statistical properties. Therefore, a multisample extension of PCH-EM can be perceived as a general purpose, reduced uncertainty, alternative to the PT method.

In this correspondence, a multisample generalization of the PCH-EM algorithm is derived and implemented.[1] To achieve this, an overview of the photon counting distribution (PCD) model, a statistical framework capable of describing data emanating from CCD, CMOS, and CMOS QIS architectures is first provided (see Section II). Subsequently, the PT method (see Section III) and the multisample PCH-EM algorithm (see Section IV) are deduced from this model, and an argument for the need of a multisample method is expounded upon (see Section V). Given that PCH-EM requires initial starting points, a statistical procedure is formulated based on initial PT estimation, subsequently refining these preliminary estimates through repeated iteration of the PCH-EM algorithm. Synthetic data from a sensor with DSERN are then subjected to parameter estimation using both the PT and PCH-EM algorithms, thereby demonstrating the enhancements achievable through PCH-EM (see Section VI). Leveraging the concept of sufficient statistics, a theoretical justification is presented to explain why PCH-EM outperforms PT, particularly within the DSERN domain, which is tied to the lossy compression PT performs when reducing raw sensor data to means and variances. Finally, experimental data captured from an early CMOS QIS sensor are used to demonstrate the algorithm's usage and validate theoretical insights (see Section VII). This work culminates with a discussion on future avenues of research and potential enhancements for the algorithm's refinement.

[1]For a full demonstration of the implemented algorithm, see the MathWorks file exchange (FEX) submission [25] and the discussion in Section VII.

II. PCD Model

The PCD model describes random excursions in a pixel's gray value X [in units of digital numbers (DNs)] according to the Poisson–Gaussian mixture model

$$X = (K + R)/g + \mu \tag{1}$$

which is used throughout the literature in applications tied to astronomy and microscopy [26], [27], [28], [29], [30]. Here, the following hold.

1) $K \sim \text{Poisson}(H)$ with quanta exposure H (e−).
2) $R \sim \mathcal{N}(0, \sigma^2)$ with read noise σ (e−).
3) g (e−/DN) is the conversion gain.
4) μ (DN) is the dc offset.

The electron number (K) represents the total number of free electrons generated per integration time from contributions of dark current and impinging photons, while the read noise (R) consists of noise introduced by the pixel's source follower and subsequent circuitry. To account for quantization, the read noise is decomposed into contributions from input-referred analog read noise (σ_{read}) and quantization noise (σ_{quan}) as $\sigma = (\sigma_{\text{read}}^2 + \sigma_{\text{quan}})^{1/2}$. In what follows, both σ_{read} and σ will be referred to as "read noise" with the distinction being that only σ can be estimated from raw gray values observed from the pixel. Given that g is assumed to be a constant, the PCD represents a linear model that does not account for conversion gain nonlinearity.

Since the conditional density of X given K is normal, i.e., $X|K \sim \mathcal{N}(\mu + K/g, (\sigma/g)^2)$, it follows that the joint distribution of the complete data (X, K) is given by

$$f_{XK}(x, k|\theta) = \frac{e^{-H} H^k}{k!} \phi\left(x; \mu + k/g, (\sigma/g)^2\right) \tag{2}$$

where $\theta = (H, g, \mu, \sigma)$ denotes the PCD parameters and $\phi(x; \alpha, \beta^2) = (1/(2\pi)^{1/2}\beta) \exp(-(x - \alpha)^2/(2\beta^2))$ denotes the Gaussian probability density with mean α and variance

β^2. The PCD is then derived by integrating the complete-data distribution across all possible states of the unobservable electron number K yielding

$$f_X(x|\theta) = \sum_{k=0}^{\infty} \frac{e^{-H} H^k}{k!} \phi\big(x; \mu + k/g, (\sigma/g)^2\big). \tag{3}$$

III. Photon Transfer

PT has been a well-established characterization methodology since the 1970s, utilized for estimating parameters of the PCD model from raw sensor data. The method involves capturing experimental data at multiple exposures, i.e., multiple H values, via one of two typical approaches: 1) maintaining a constant source flux while varying integration time or 2) keeping integration time constant while adjusting source flux. While there are many versions of the PT method, all of these versions share in common a data reduction strategy involving the reduction of raw data to sample means and variances followed by regression procedures to estimate the parameters. For the sake of brevity, only a constant flux version of the technique will be derived here.

Suppose Φ (e$-$/s) is the electron flux and τ (s) is the integration time. For this derivation, the flux will be assumed to be independent of τ. In this manner, Φ represents the rate at which electrons are freed, so that the quanta exposure observed by the pixel is $H = \Phi\tau$. Note that if no light source is used in the experiment, then Φ can be directly interpreted as the pixel's dark current. Following (1), the mean and variance of the pixel output can be parameterized in terms of integration time as $(\mathsf{E}X)(\tau) = \Phi\tau/g + \mu$ and $(\mathsf{Var}X)(\tau) = \Phi\tau/g^2 + (\sigma/g)^2$, respectively.

To perform the constant flux PT experiment, consider a set of J integration times $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_J)$, each with a corresponding random variable $X_j \sim \text{PCD}(\Phi\tau_j, g, \mu, \sigma)$. From X_j, N_j i.i.d. observations of the pixel are drawn, resulting in the random sample $\mathbf{x}_j = (x_{j,1}, \ldots, x_{j,N_j})$. Random samples captured at all J integration times combined represent the multisample dataset $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_J)$. Since the integration times are known, this effectively renders the multisample dataset with four unknown parameters to be estimated: $\theta = (\Phi, g, \mu, \sigma)$.

Once data are captured, the estimation procedure begins by reducing the multisample data to means and variances resulting in $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_J)$ and $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_J)$, where $\bar{x}_j = (1/N_j)\sum_{n=1}^{N_j} x_{jn}$ and $\hat{x}_j = (1/N_j)\sum_{n=1}^{N_j}(x_{jn} - \bar{x}_j)^2$ are the sample mean and variance of $\mathbf{x}_j$, respectively. It is noted that throughout this work, bar notation $\bar{(\cdot)}$ is used to denote means, while hat notation $\hat{(\cdot)}$ denotes variances and covariances.

Following data reduction, least squares regression procedures are used to extract the linear relationship in the data $(\boldsymbol{\tau}, \bar{\mathbf{x}})$, yielding $\bar{x}(\tau) = m_1\tau + b_1 \approx (\Phi/g)\tau + \mu$. Similarly, least squares regression is again used to extract the linear relationship in the data $(\boldsymbol{\tau}, \hat{\mathbf{x}})$, resulting in $\hat{x}(\tau) = m_2\tau + b_2 \approx (\Phi/g^2)\tau + (\sigma/g)^2$. With the slopes and intercepts of these linear fits, the PCD parameters for the pixel are estimated via

$$\begin{aligned} \tilde{\Phi} &= \tilde{g} m_1, \quad \tilde{g} = m_1/m_2 \\ \tilde{\mu} &= b_1, \quad \tilde{\sigma} = \tilde{g}\sqrt{b_2}. \end{aligned} \tag{4}$$

IV. Multisample PCH-EM

In the PT method, raw multisample data were reduced to means and variances, and the linear relationships between the means and variances were exploited to characterize the pixel under test. In contrast to this approach, the PCH-EM algorithm performs MLE on the same multisample data by reducing it to a log-likelihood function

$$\ell(\theta|\mathbf{x}) = \sum_{j=1}^{J}\sum_{n=1}^{N_j} \log \sum_{k=0}^{\infty} \frac{e^{-H_j} H_j^k}{k!} \phi\big(x_{jn}; \mu + k/g, (\sigma/g)^2\big) \tag{5}$$

and then determining the parameters that maximize ℓ.

As is common with many MLE problems, an explicit expression for the MLEs $\tilde{\theta} = \arg\max_\theta \ell(\theta|\mathbf{x})$ is intractable. Numerical methods, such as the Newton–Raphson iteration, may be used to numerically maximize the log likelihood; however, poor initial guesses can lead to numerical instability when inverting the Hessian matrix, so this approach can be undesirable [31].

A popular, more stable, alternative for maximizing log-likelihood functions is the expectation–maximization (EM) algorithm, which instead of maximizing the log likelihood directly, iteratively maximizes a related, often simpler, function to compute MLEs [32]. To see how, note that through an application of Bayes' theorem, one may decompose the log likelihood (5) into $\ell(\theta|\mathbf{x}) = Q(\theta|\theta^{(t)}) + H(\theta|\theta^{(t)})$, where $Q(\theta|\theta^{(t)}) = \mathsf{E}_{\theta^{(t)}}(\log f_{XK}(\mathbf{x}, \mathbf{k}|\theta))$ is the *expected complete-data log likelihood*. Here, the expectation is taken with respect to the conditional distribution $p_{K|X}(\mathbf{k}|\mathbf{x}, \theta^{(t)})$, and $\theta^{(t)}$ is a guess of the PCD parameters. Defining $\theta^{(t+1)} = \arg\max_\theta Q(\theta|\theta^{(t)})$, one can guarantee

$$\ell\big(\theta^{(t+1)}\big|\mathbf{x}\big) - \ell\big(\theta^{(t)}\big|\mathbf{x}\big) \geq Q\big(\theta^{(t+1)}\big|\theta^{(t)}\big) - Q\big(\theta^{(t)}\big|\theta^{(t)}\big) \tag{6}$$

showing that the updated parameter $\theta^{(t+1)}$ has a log likelihood greater than or equal to the initial guess (in practice, the increase can be substantial). By repeating the process of maximizing Q, each time redefining it in terms of the current parameter estimate, one can monotonically increase sample likelihood with each iteration. The benefit of doing this in the specific use case of PCH-EM is that the complete-data distribution (2) forms an exponential family; thus, Q can be maximized in closed form, resulting in a set of update equations that monotonically "climb" the log-likelihood function without having to evaluate derivatives or perform matrix inversions.

A. Multisample PCH-EM Update Equations

Similar to PT, multisample PCH-EM data capture can be performed in either a constant flux or constant exposure mode. Combinations of these approaches can also be accommodated by specifying the form of each H_j in the Q function. Note that the ability to modify Q to reflect the specifics of the experiment performed shows that multisample PCH-EM should be considered a family of algorithms and not one single set of update equations. Mimicking the previous section, a strictly constant flux version of the algorithm will be derived here.

It is noted that a similar procedure can be followed to derive the constant exposure method.

Begin with the same multisample data captured in PT, and let $(\mathbf{p}_j, \mathbf{b}_j) = ((p_{j,1}, b_{j,1}), \ldots, (p_{j,\#_j}, b_{j,\#_j}))$ denote the histogram of the jth sample $\mathbf{x}_j$. Here, $\mathbf{b}_j$ is the vector of unique gray values in $\mathbf{x}_j$ and $\mathbf{p}_j$ that are sample probabilities (counts in each bin b_{jn} normalized by N_j). Using this convention, $\#_j$ represents the number of unique gray values in the jth sample. Using these histograms and the definition of Q, one is able to derive

$$Q\left(\theta\middle|\theta^{(t)}\right) = \sum_{j=1}^{J} N_j \sum_{n=1}^{\#_j} p_{jn} \times \sum_{k=0}^{\infty} \gamma_{jnk}^{(t)} \left(-\Phi\tau_j + k \log \Phi\tau_j - \log(\sigma/g) - \frac{\left(b_{jn} - \mu - k/g\right)^2}{2(\sigma/g)^2} + C \right) \quad (7)$$

where C is a constant independent of θ and

$$\gamma_{jnk}^{(t)} = \left. \frac{\frac{e^{-\Phi\tau_j}(\Phi\tau_j)^k}{k!}\phi\left(b_{jn}; \mu + k/g, (\sigma/g)^2\right)}{\sum_{\ell=0}^{\infty} \frac{e^{-\Phi\tau_j}(\Phi\tau_j)^\ell}{\ell!}\phi\left(b_{jn}; \mu + \ell/g, (\sigma/g)^2\right)} \right|_{\theta=\theta^{(t)}}. \quad (8)$$

Here, the *membership probabilities* $\gamma_{jnk}^{(t)}$ represent the probability that the electron number K equals k given the observed gray value x_{jn} and the current parameter estimate $\theta^{(t)}$. Q is then maximized by solving for the critical point $\nabla_\theta Q = 0$, yielding the update equations (see [33] for derivation)

$$\Phi^{(t+1)} = \bar{k}^{(t)} / \bar{\tau} \quad (9a)$$

$$g^{(t+1)} = \hat{k}^{(t)} / \widehat{xk}^{(t)} \quad (9b)$$

$$\mu^{(t+1)} = \bar{x} - \bar{k}^{(t)} / g^{(t+1)} \quad (9c)$$

$$\sigma^{(t+1)} = \left(\hat{x}\left(g^{(t+1)}\right)^2 - \hat{k}^{(t)} \right)^{1/2} \quad (9d)$$

where $\bar{\tau} = \sum_{j=1}^{J} w_j \tau_j$ with $w_j = (N_j / (\sum_{m=1}^{J} N_m))$ and

$$\bar{x} = \sum_{j=1}^{J} w_j \sum_{n=1}^{\#_j} p_{jn} b_{jn} \quad (10a)$$

$$\hat{x} = \sum_{j=1}^{J} w_j \sum_{n=1}^{\#_j} p_{jn} \left(b_{jn} - \bar{x}\right)^2 \quad (10b)$$

$$\bar{k}^{(t)} = \sum_{j=1}^{J} w_j \sum_{n=1}^{\#_j} p_{jn} \sum_{k=0}^{\infty} \gamma_{jnk}^{(t)} k \quad (10c)$$

$$\hat{k}^{(t)} = \sum_{j=1}^{J} w_j \sum_{n=1}^{\#_j} p_{jn} \sum_{k=0}^{\infty} \gamma_{jnk}^{(t)} \left(k - \bar{k}^{(t)}\right)^2 \quad (10d)$$

$$\widehat{xk}^{(t)} = \sum_{j=1}^{J} w_j \sum_{n=1}^{\#_j} p_{jn} \left(b_{jn} - \bar{x}\right) \sum_{k=0}^{\infty} \gamma_{jnk}^{(t)} \left(k - \bar{k}^{(t)}\right). \quad (10e)$$

Supplied with an initial guess/estimate of the parameters $\theta^{(0)} = (\Phi^{(0)}, g^{(0)}, \mu^{(0)}, \sigma^{(0)})$, the multisample PCH-EM algorithm iteratively updates (improves) the initial guess by the following: 1) calculating the statistics in (10) (E-step) and 2) updating the estimates with the update equations (9) (M-step). These two steps are repeated until convergence to a fixed point is achieved, which is usually determined when the change in log likelihood is negligible, i.e., $\ell(\theta^{(t+1)}|\mathbf{x}) - \ell(\theta^{(t)}|\mathbf{x}) \leq \epsilon$ for some appropriate choice of ϵ.

For large read noise ($\sigma_{\text{read}} \gg 1$ e$-$), the individual Gaussian components of the PCD exhibit significant overlap leading to slower convergence of the algorithm [34]. Because of this, it is generally advisable for practical implementations of PCH-EM to also include a maximum allowable number of iterations that overrides the log-likelihood criteria when it cannot be achieved within a reasonable amount of time.

V. Why Adopt a Multisample Approach?

When a pixel exhibits DSERN, the raw data it produces contain sufficient information to permit estimating the PCD parameters from a single sample. This assertion is supported by the fact that reasonable parameter estimates can be obtained simply by visually examining the histogram generated by a DSERN pixel. However, as the read noise increases, the information content of the data decreases, and its distribution tends toward that of the normal distribution. In this high noise regime, the minimal sufficient statistic of the data consists of a sample mean and variance for each sample captured. For a single-sample approach, this implies that the data are completely summarized by a 2-D statistic; however, the constant flux model has four parameters [$\theta = (\Phi, g, \mu, \sigma)$]; thus, a single sample does not provide enough information about the parameters for reliable estimation in high noise conditions. This limitation applies not only to single-sample PCH-EM but also to other single-sample methods, such as the photon counting histogram (PCH) method, valley peak modulation (VPM) method, peak separation and width (PSW) method, and FFT method [19], [20], [23].

By adding an additional sample captured at a different integration time, the minimal sufficient statistic in high noise conditions becomes 4-D, matching the dimension of the unknown parameter vector implying enough information has been captured to ensure reliable estimation even at high read noise. Additional samples further serve to ensure enough information is captured. This is particularly crucial as modern CMOS QISs exhibit a distribution of read noise across all pixels in the sensor array. While most pixels exhibit DSERN, there is typically a subset of pixels with read noise outside the DSERN regime for which single-sample methods produce worse estimates compared with PT [19]. By adopting a multisample approach, the algorithm becomes robust to higher read noise and can accurately characterize all pixels.

Combining multiple samples into a single estimation procedure not only enhances the algorithm's resilience against high read noise but also facilitates obtaining more accurate parameter estimates. To see why, note that if the sample sizes are large and the read noise is small, the determinant of the PCH-EM estimate's covariance matrix (generalized variance) can be approximated by $\det \Sigma \sim \sigma^6/(2N^4\,\bar{\tau}(\bar{\tau} + \Phi\hat{\tau}))$, with $N = \sum_{j=1}^{J} N_j$ denoting the total sample size and $\hat{\tau} = \sum_{j=1}^{J} w_j(\tau_j - \bar{\tau})^2$ the variance of the integration times [33].

TABLE I
SIMULATION PARAMETERS

parameter	value	unit
Φ	5	$e\text{-}/s$
g	$\sigma_{\text{read}}/4$	$e\text{-}/\text{DN}$
μ	100	DN
σ_{read}	$\text{linspace}(0.05, 1, 64)$	$e\text{-}$
τ_1	0.01	s
τ_2	1	s

Now, consider a two-sample procedure where the total sample size N is fixed, so that $N_1 = (1-p)N$ and $N_2 = pN$ for some $p \in [0, 1]$ and suppose $\Phi = 10$ e−/s, $\sigma = 0.15$ e−, $\tau_1 = 0.1$ s, and $\tau_2 = 5$ s. At $p = 0$, the generalized variance corresponds to that of a single-sample method with integration time τ_1. Likewise, at $p = 1$, it corresponds to that of a single-sample method with integration time τ_2. Calculating the approximation for $\det \Sigma$ for several p values reveals that it is minimized when $p \approx 0.7$, demonstrating that a two-sample approach can produce lower variance compared with a one-sample approach, even with the same total number of observations. This phenomenon can be attributed to "information diversification." Because the amount of information about each parameter present in the raw sensor data varies dynamically with quanta exposure $H = \Phi\tau$, by combining multiple samples with varying H, the collective information in the multisample data is diversified across all parameters, generally leading to improved overall uncertainty in the parameter estimates.

VI. MONTE CARLO DEMONSTRATION

Monte Carlo experiments were performed to demonstrate the decrease of estimate uncertainty achievable with PCH-EM. For the experiment, a two-sample ($J = 2$) data capture was adopted, and a pixel with the parameters outlined in Table I was specified. A total of 64 linearly space values of the read noise on the interval [0.05, 1] were chosen to help observe the effect read noise has on estimate uncertainty and the ability of PCH-EM to improve PT estimates.

An important consideration in the experimental design is the sample sizes. As read noise increases, the uncertainty in the estimates of both methods will also increase [35], [36]. To keep the uncertainty of the estimates from growing too large, the sample sizes were determined via [37]

$$N_1 = \left\lceil \frac{2\zeta(1+\zeta)}{\delta^2(1-\zeta)^2} + 1 \right\rceil \quad \text{and} \quad N_2 = \left\lceil \frac{2(1+\zeta)}{\delta^2(1-\zeta)^2} + 5 \right\rceil \tag{11}$$

where $\delta = 0.02$ and $\zeta = (\Phi\tau_1 + \sigma_{\text{read}}^2)/(\Phi\tau_2 + \sigma_{\text{read}}^2)$. These sample sizes will increase with increasing read noise and subsequently force the relative uncertainty of the PT conversion gain estimate in (4) to be approximately equal to δ (2% error), independent of the parameter values.

The experiment was performed by generating synthetic data, estimating the PCD parameters with PT, and then refining the PT estimates with PCH-EM. This experiment was repeated a total of $M = 1000$ times at each of the 64 parameter values. To measure each method's estimation error, a relative root-mean-squared error (RRMSE) metric

$$\text{RRMSE}(\tilde{\theta}_i) = \left(\frac{1}{M} \sum_{n=1}^{M} \left(\frac{\tilde{\theta}_{in} - \theta_i}{\theta_i} \right)^2 \right)^{1/2} \approx \frac{\sqrt{\mathsf{MSE}(\tilde{\theta}_i)}}{|\theta_i|} \tag{12}$$

was used. Here, θ_i is replaced with one of Φ, g, μ, or σ; $\tilde{\theta}_{in}$ is the nth estimate of θ_i from either PT or PCH-EM; and $\mathsf{MSE}(\tilde{\theta}_i) = \mathsf{E}(\tilde{\theta}_i - \theta_i)^2$ is the exact value of the estimator mean-squared error (MSE).

For the sample sizes used, the PCH-EM estimates should be nearly unbiased, so that one can further approximate $\text{RRMSE}(\tilde{\theta}_i) \approx (\mathsf{Var}(\tilde{\theta}_i))^{1/2}/|\theta_i|$. Calculating the Fisher information of the multisample data $I_{\mathbf{X}}(\theta) = -\mathsf{E}(\nabla_\theta \nabla_\theta^T \ell(\theta|\mathbf{X}))$, the Cramér–Rao lower bound (CRLB) for the RRMSE can be calculated via $((I_{\mathbf{X}}^{-1}(\theta))_{ii})^{1/2}/|\theta_i| \leq (\mathsf{Var}(\tilde{\theta}_i))^{1/2}/|\theta_i|$, and this bound can be compared with the RRMSE of both PT and PCH-EM.

Fig. 2 shows the RRMSE and CRLB curves obtained from the experiment as a function of read noise. As can be observed, the relative uncertainty of the PCH-EM estimates is less than those of PT at all read noise values, particularly in the DSERN regime. For example, at $\sigma_{\text{read}} = 0.214$ e−, the RRMSE of the PT estimate for Φ was reduced via PCH-EM by 73%, while the RRMSE of the PT estimates for g, μ, and σ was reduced by 93%, 74%, and 96%, respectively. The fact that the PCH-EM curves are nearly identical to the CRLB curves indicates that for the specified parameters, PCH-EM is producing the best possible estimates. In addition, for all four parameters, the PCH-EM RRMSE curves approached those of PT, from below, asymptotically, as the read noise increases. The reason for this phenomenon is tied to the fact that reducing the multisample data to a likelihood function is a minimal sufficient encoding, incurring no information loss, whereas the means and variances of the PT method are not sufficient, resulting in information loss and higher estimate uncertainty (for a detailed discussion, see [33, Sec. VII]). However, as read noise increases, that raw data increasing becomes normally distributed where the means and variances are sufficient statistics causing the information loss incurred in PT to vanish, resulting in both methods achieving similar uncertainties [33]. This characteristic of PCH-EM can, thus, be viewed as an overall improvement over the PT method, since it achieves less uncertainty than PT everywhere and is asymptotically equivalent to PT at large read noise, bridging the gap between single-sample methods that improve PT only in the DSERN regime while also being able to characterize at a continuum of read noise levels, such as PT.

From a historical perspective, the fact that PCH-EM only begins to gain significant advantage over PT in the DSERN regime gives context to why such a method is not already in widespread use for sensor characterization. Had PCH-EM been developed to characterize early CCDs in the 1970s, it would have exhibited little to no improvement over the much simpler and computationally friendly PT method, rendering PT the preferred technique. With the advent of DSERN technology,

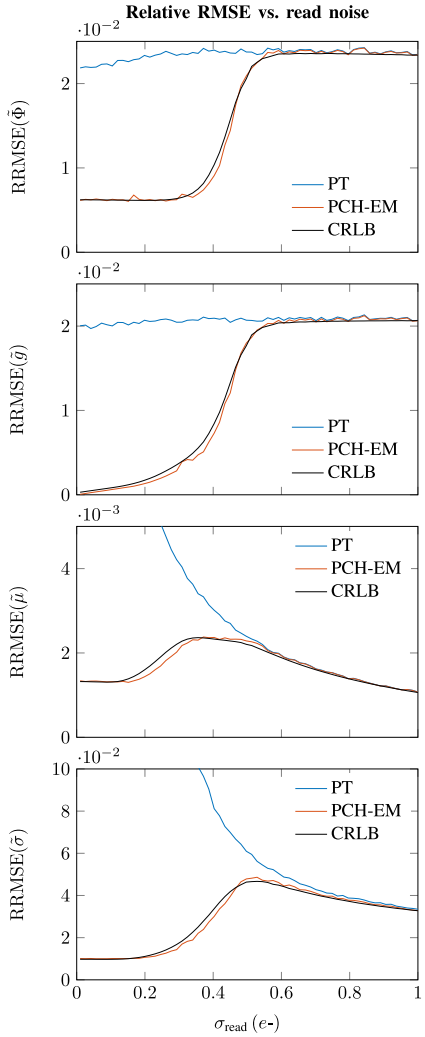

Fig. 2. RRMSE versus read noise for parameter estimates computed using constant flux implementation of PT and PCH-EM. RRMSE curves for PT $\tilde{\mu}$ and $\tilde{\sigma}$ grow large near $\sigma_{\text{read}} = 0$ and were clipped from the plot window.

PCH-EM can now deliver superior estimates over PT, thus establishing its utility.

VII. Experimental Demonstration

Experimental constant flux data were captured with the Gigajot Technology Inc. (GJ00111) "Cleveland" QIS camera. The constant source of electron flux came from a combination of the sensor's internal dark current and impinging photons from an incandescent light source coupled with an integrating sphere to provide the focal plane with a stable, flat field. For the experiment, a three-sample approach was adopted with integration times and sample sizes given in Table II. As is the case in Table II, the sample sizes should be on the order of thousands to tens of thousands and increasing with quanta exposure. More sophisticated allocation of sample sizes can be accomplished by choosing the sample sizes that maximize the determinant of the multisample Fisher information $\det \boldsymbol{I}_{\mathbf{X}}(\theta)$.

Experimental data were processed in MATLAB using the code uploaded to the MathWorks FEX [25]. It is noted that the code available on the FEX provides a full demonstration of

TABLE II
Experimental Parameters

	int. time (τ_j)	sample size (N_j)
sample 1	0.876 ms	8500
sample 2	3.3 ms	14000
sample 3	6.6 ms	20000

TABLE III
Small Noise Pixel Characterization (See Fig. 3)

	PT	PCH-EM	std. error (PCH-EM)
$\tilde{\Phi}$ (e-/s)	3175	3154	± 17
$\tilde{g}$ (e-/DN)	0.0862859140	0.0841397938	$\pm 7.3 \times 10^{-10}$
$\tilde{\mu}$ (DN)	2.955	-0.070	$\pm 3.7 \times 10^{-3}$
$\tilde{\sigma}$ (e-)	0.667807	0.286082	$\pm 2.1 \times 10^{-6}$

constant flux PCH-EM by generating synthetic data. Here, the code was modified by removing the synthetic data generation, instead accepting raw experimental data from the Cleveland camera experiment. Minimal effort was put into optimizing the efficiency of the code with the intent to make it more readable and track with the equations laid out in this work.

The code follows a step-by-step procedure in the main file `demo.m` to processes the raw multisample data by first organizing it into the data structure `Struct` (`fluxData.m`) and then passing this structure to various steps in data processing pipeline. In `fluxData.m`, the data for each sample are reduced to a histogram, and these histograms, along with the raw samples, are used throughout the processing pipeline. The data processing pipeline following `fluxData.m` consists of the following: 1) performing initial starting point estimation (`fluxStart.m`); 2) performing PCH-EM estimation seeded with the starting points (`fluxPCHEM.m`); and 3) computing the covariance matrix for the final PCH-EM estimates (`fluxCovariance.m`).

The function `fluxStart.m` uses a hybrid approach to starting point estimation by estimating the starting points with the FFT method (`fluxFFT.m`), described in [22] and [23], if only a single sample is captured or the constant flux PT method (`fluxPT.m`) if two or more samples are captured. By using this hybrid approach, the code works for any numbers of samples. Likewise, `fluxCovariance.m` estimates the PCH-EM covariance matrix by computing the observed information matrix $J(\tilde{\theta}) = -\nabla_\theta \nabla_\theta^T \ell(\theta|\mathbf{x})|_{\theta=\tilde{\theta}}$, where $\tilde{\theta}$ is the PCH-EM estimate of the parameters. In this way, J^{-1} becomes an estimate for the PCH-EM estimate covariance matrix, and the diagonal elements of J^{-1} represent standard error estimates for each PCH-EM parameter estimate. Additional functions are also supplied in the FEX submission to display the final estimates and provide plots pertaining to algorithm convergence and data visualization.

Tables III and IV present the characterization results for two of the QIS pixels tested: one with small read noise and another with larger read noise. It is important to note that the camera software has built-in offset subtraction, which is why $\tilde{\mu}$ in Table III can have a negative value in the absence of data truncation. Comparing the standard errors in

TABLE IV
LARGER NOISE PIXEL CHARACTERIZATION (SEE FIG. 4)

	PT	PCH-EM	std. error (PCH-EM)
$\tilde{\Phi}$ (e-/s)	3117	3074	± 820
$\tilde{g}$ (e-/DN)	0.0852220	0.0840197	$\pm 5.6 \times 10^{-7}$
$\tilde{\mu}$ (DN)	8.18	8.16	$\pm 7.1 \times 10^{-2}$
$\tilde{\sigma}$ (e-)	0.909	0.792	$\pm 1.0 \times 10^{-3}$

both tables not only reveals the impact read noise has on estimate uncertainty but also the level of precision obtainable at low read noise. For example, the PCH-EM standard error estimate for $\tilde{g}$ in Table III indicates approximately ten decimal places of precision. This error estimate was independently calculated through Monte Carlo experiments, verifying that the observed information matrix gave accurate uncertainty estimates. Corresponding to these tables are Figs. 3 and 4 showing the log likelihood obtained at each iteration of PCH-EM (top) and quality of fit obtained with PT and PCH-EM (bottom). To visualize the quality of fit, the multisample data for each pixel were turned into a "stacked" histogram with gray shade indicating the frequency in each bin obtained from each sample. This stacked histogram was then overlaid with the estimated multisample PCD: $f(x|\tilde{\theta}) = \sum_{j=1}^{J} w_j f_X(x|\tilde{\theta}_j)$, where f_X is the PCD in (3), and $\tilde{\theta}_j = (\tilde{\Phi}\tau_j, \tilde{g}, \tilde{\mu}, \tilde{\sigma})$ are the PT or PCH-EM parameter estimates.

In reference to the data in Fig. 3, one can observe that the algorithm monotonically increased log likelihood (improved the PT estimates) with each iteration, and the fit obtained with PT (red curve) was substantially improved with PCH-EM (blue curve). Likewise, now referencing the data in Fig. 4, one can observe that the noise is large enough to yield approximately normally distributed data, and that the PCH-EM fit is nearly identical to the PT fit as is to be expected as larger read noise. As was pointed out in Section IV-A, at larger noise, the peaks in the PCD exhibit more overlap causing the algorithm convergence to decrease, and this phenomenon is reflected by the larger number of iterations PCH-EM needed to converge in the large noise pixel data. Despite this slower convergence, the algorithm was still able to monotonically increase log likelihood at each iteration. The ability of multisample PCH-EM to handle both small and large noise pixels validates its usage as a general characterization algorithm not limited to pixels with DSERN.

VIII. CONCLUSION

Camera companies and camera consumers alike rely on characterization techniques to troubleshoot, optimize, and calibrate image sensors. As sensor technology continues to evolve, new characterization methods are developed to measure key performance parameters and inform decision-making. In this correspondence, the next step in this evolution was taken by developing and implementing a multisample PCH-EM algorithm, which shows dramatic improvement in estimate uncertainty over the PT method, particularly in the regime of DSERN. It was shown that the source of this uncertainty advantage is tied to the lossy compression PT performs on the raw sensor data by reducing it to sample means and variances.

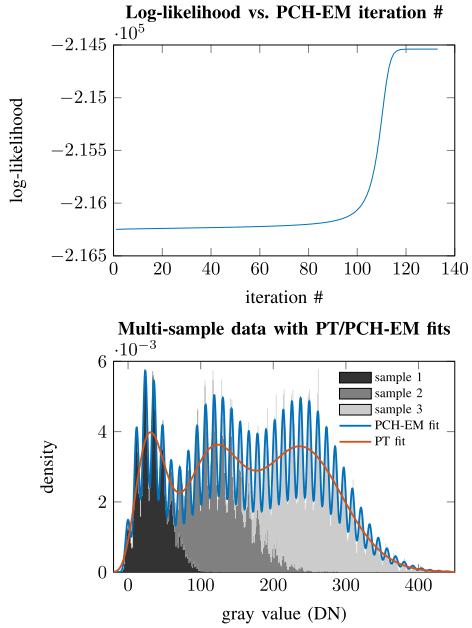

Fig. 3. Log likelihood versus PCH-EM iteration number (top) and experimental multisample data with PT/PCH-EM fits (bottom) for the small noise pixel in Table III.

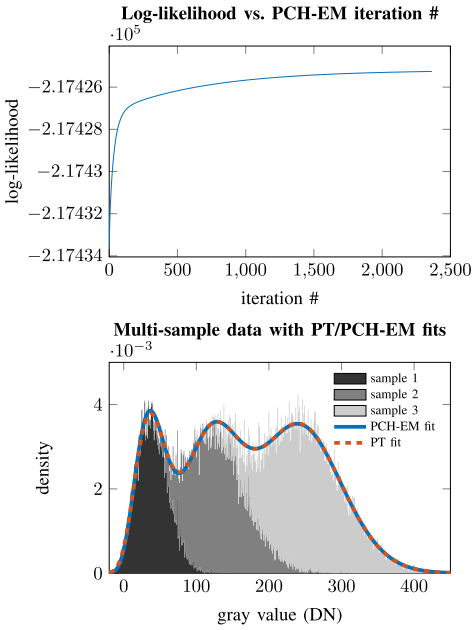

Fig. 4. Log likelihood versus PCH-EM iteration number (top) and experimental multisample data with PT/PCH-EM fits (bottom) for the larger noise pixel in Table IV.

As read noise exceeds the DSERN threshold, the information loss incurred by PT vanishes, and both methods show nearly identical estimate uncertainty; thus, PCH-EM can be viewed as a general improvement over the PT method.

This maximum likelihood approach to sensor characterization, implemented via EM, is in actuality a general approach not limited to a single algorithm or single sensor technology. By constructing various Q functions reflecting an assumed sensor model and experimental parameters, a large number of PCH-EM algorithms can be derived. Here, the process of Q-function construction and update equation derivation was

demonstrated for a single pixel, constant flux, approach to PCH-EM using multiples samples. However, by combining data from multiple pixels, other parameters, such as photoresponse nonuniformity (PRNU), can be incorporated into the Q function, expanding the measurement capabilities of the method. Other opportunities for improving the reliability of the algorithm also exist. It is understood that PCH-EM is a local optimization strategy prone to converging to a nonoptimal local maxima of the likelihood function if provided with poor starting points. As such, future investigations will look into not only expanding the measurement capabilities of the PCH-EM approach but also relaxing the algorithm's dependence on starting point location, thus providing a more comprehensive methodology to sensor characterization.

REFERENCES

[1] D. M. Hunten and C. J. Stump, "Performance of a silicon Vidicon at low signal levels," *Appl. Opt.*, vol. 15, no. 12, p. 3105, Dec. 1976.

[2] J. R. Janesick, *Photon Transfer: DN → λ*. Bellingham, WA, USA: SPIE, 2007.

[3] J. R. Janesick, *Personal Communication*, Jan. 2024.

[4] M. F. Tompsett et al., "Charge-coupled imaging devices: Experimental results," *IEEE Trans. Electron Devices*, vol. ED-18, no. 11, pp. 992–996, Nov. 1971.

[5] C. H. Sequin et al., "A charge-coupled area image sensor and frame store," *IEEE Trans. Electron Devices*, vol. ED-20, no. 3, pp. 244–252, Mar. 1973.

[6] E. R. Fossum, N. Teranishi, and A. J. P. Theuwissen, "Digital image sensor evolution and new frontiers," *Annu. Rev. Vis. Sci.*, vol. 10, Oct. 2024.

[7] *EMVA Standard 1288: Standard for Characterization of Image Sensors and Cameras, Release 4.0 Linear*, EMVA 1288 Working Group, Heidelberg, Germany, Jun. 2021.

[8] J. Nakamura, *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton, FL, USA: CRC Press, 2005.

[9] E. Fossum, J. Ma, S. Masoodian, L. Anzagira, and R. Zizza, "The quanta image sensor: Every photon counts," *Sensors*, vol. 16, no. 8, p. 1260, Aug. 2016.

[10] J. Ma and E. R. Fossum, "Quanta image sensor jot with sub 0.3e- r.m.s. read noise and photon counting capability," *IEEE Electron Device Lett.*, vol. 36, no. 9, pp. 926–928, Sep. 2015.

[11] J. Ma, D. Starkey, A. Rao, K. Odame, and E. R. Fossum, "Characterization of quanta image sensor pump-gate jots with deep sub-electron read noise," *IEEE J. Electron Devices Soc.*, vol. 3, no. 6, pp. 472–480, Nov. 2015.

[12] M.-W. Seo, S. Kawahito, K. Kagawa, and K. Yasutomi, "A 0.27e-rms read noise 220-μV/e-conversion gain reset-gate-less CMOS image sensor with 0.11-μm CIS process," *IEEE Electron Device Lett.*, vol. 36, no. 12, pp. 1344–1347, Dec. 2015.

[13] S. Masoodian, A. Rao, J. Ma, K. Odame, and E. R. Fossum, "A 2.5 pJ/b binary image sensor as a pathfinder for quanta image sensors," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 100–105, Jan. 2016.

[14] J. Ma, S. Chan, and E. R. Fossum, "Review of quanta image sensors for ultralow-light imaging," *IEEE Trans. Electron Devices*, vol. 69, no. 6, pp. 2824–2839, Jun. 2022.

[15] J. Tiffenberg et al., "Single-electron and single-photon sensitivity with a silicon skipper CCD," *Phys. Rev. Lett.*, vol. 119, no. 13, Sep. 2017, Art. no. 131802.

[16] K. Nakamoto and H. Hotaka, "Efficient and accurate conversion-gain estimation of a photon-counting image sensor based on the maximum likelihood estimation," *Opt. Exp.*, vol. 30, no. 21, pp. 37493–37506, Oct. 2022.

[17] B. A. Cervantes-Vergara et al., "Skipper-CCDs: Current applications and future," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 1046, Jan. 2023, Art. no. 167681.

[18] A. J. Lapi et al., "Skipper-in-CMOS: Non-destructive readout with sub-electron noise performance for pixel detectors," 2024, *arXiv:2402.12516*.

[19] D. A. Starkey and E. R. Fossum, "Determining conversion gain and read noise using a photon-counting histogram method for deep sub-electron read noise image sensors," *IEEE J. Electron Devices Soc.*, vol. 4, no. 3, pp. 129–135, May 2016.

[20] N. Dutton, I. Gyongy, L. Parmesan, and R. Henderson, "Single photon counting performance and noise analysis of CMOS SPAD-based image sensors," *Sensors*, vol. 16, no. 7, p. 1122, Jul. 2016.

[21] A. J. Lapi, F. Chierchie, and G. F. Moroni, "Gain calibration and nonlinearity analysis in single photon sensitivity skipper CCD," in *Proc. 21st Workshop Inf. Process. Control (RPIC)*, Nov. 2021, pp. 1–6.

[22] A. J. Hendrickson and D. P. Haefner, "Photon counting histogram expectation maximization algorithm for characterization of deep sub-electron read noise sensors," *IEEE J. Electron Devices Soc.*, vol. 11, pp. 367–375, 2023.

[23] A. Hendrickson and D. P. Haefner, "A comparative study of methods to estimate conversion gain in sub-electron and multi-electron read noise regimes," *Proc. SPIE*, vol. 12533, Sep. 2023, Art. no. 125330R.

[24] A. J. Hendrickson, D. P. Haefner, N. R. Shade, and E. R. Fossum, "Experimental verification of PCH-EM algorithm for characterizing DSERN image sensors," *IEEE J. Electron Devices Soc.*, vol. 11, pp. 376–384, 2023.

[25] A. J. Hendrickson, D. P. Haefner, S. H. Chan, N. R. Shade, and E. R. Fossum. (2024). *Multi-Sample PCH-EM Algorithm*. MATLAB Central File Exchange. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/158931-multi-sample-pch-em-algorithm

[26] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser, "Multiframe sure-let denoising of timelapse fluorescence microscopy images," in *Proc. 5th IEEE Int. Symp. Biomed. Imag., From Nano Macro*, May 2008, pp. 149–152.

[27] B. Begovic, V. Stankovic, and L. Stankovic, "Contrast enhancement and denoising of Poisson and Gaussian mixture noise for solar images," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 185–188.

[28] G. Fatima and P. Babu, "PGPAL: A monotonic iterative algorithm for phase-retrieval under the presence of Poisson-Gaussian noise," *IEEE Signal Process. Lett.*, vol. 29, pp. 533–537, 2022.

[29] D. A. Barmherzig and M. Eickenberg, "Low-photon holographic phase retrieval with Poisson-Gaussian denoising," in *Proc. Imag. Appl. Opt. Congr. (3D, AOA, COSI, ISA, pcAOP)*, Sep. 2022.

[30] J. Hu, Z. Li, X. Xu, L. Shen, and J. A. Fessler, "Poisson-Gaussian holographic phase retrieval with score-based image prior," in *Proc. NeurIPS Workshop Deep Learn. Inverse Problems*, 2023.

[31] M. R. Gupta, "Theory and use of the EM algorithm," *Found. Trends Signal Process.*, vol. 4, no. 3, pp. 223–296, Apr. 2011.

[32] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.

[33] A. J. Hendrickson, D. P. Haefner, S. H. Chan, N. R. Shade, and E. R. Fossum, "PCH-EM: A solution to information loss in the photon transfer method," 2024, *arXiv:2403.04498*.

[34] I. Naim and D. Gildea, "Convergence of the EM algorithm for Gaussian mixtures with unbalanced mixing coefficients," 2012, *arXiv:1206.6427*.

[35] B. P. Beecken and E. R. Fossum, "Determination of the conversion gain and the accuracy of its measurement for detector elements and arrays," *Appl. Opt.*, vol. 35, no. 19, pp. 3471–3477, Jul. 1996.

[36] A. J. Hendrickson, "Centralized inverse-fano distribution for controlling conversion gain measurement accuracy of detector elements," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 34, no. 8, p. 1411, Aug. 2017.

[37] A. Hendrickson, D. P. Haefner, and B. L. Preece, "On the optimal measurement of conversion gain in the presence of dark noise," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 39, no. 12, pp. 2169–2185, Dec. 2022.