

# Fisher information matrix for branching processes with application to electron-multiplying charge-coupled devices

Jerry Chao · E. Sally Ward · Raimund J. Ober

Received: 14 September 2010 / Revised: 17 March 2011 / Accepted: 22 March 2011 /  
Published online: 8 April 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** The high quantum efficiency of the charge-coupled device (CCD) has rendered it the imaging technology of choice in diverse applications. However, under extremely low light conditions where few photons are detected from the imaged object, the CCD becomes unsuitable as its readout noise can easily overwhelm the weak signal. An intended solution to this problem is the electron-multiplying charge-coupled device (EMCCD), which stochastically amplifies the acquired signal to drown out the readout noise. Here, we develop the theory for calculating the Fisher information content of the amplified signal, which is modeled as the output of a branching process. Specifically, Fisher information expressions are obtained for a general and a geometric model of amplification, as well as for two approximations of the amplified signal. All expressions pertain to the important scenario of a Poisson-distributed initial signal, which is characteristic of physical processes such as photon detection. To facilitate the investigation of different data models, a “noise coefficient” is introduced which allows the analysis and comparison of Fisher information via a scalar quantity. We apply our results to the problem of estimating the location of a point source from its image, as observed through an optical microscope and detected by an EMCCD.

**Keywords** Branching process · Electron Multiplication · Fisher information · Quantum-limited imaging · Single molecule microscopy

---

J. Chao (✉) · R. J. Ober  
Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA  
e-mail: jcsy@utdallas.edu

R. J. Ober  
e-mail: ober@utdallas.edu

J. Chao · E. S. Ward · R. J. Ober  
Department of Immunology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA  
e-mail: sally.ward@utsouthwestern.edu

## 1 Introduction

The charge-coupled device (CCD) is an important digital imaging technology that has found utility in many applications. It plays crucial roles, for example, in areas as disparate as high sensitivity cellular microscopy and astronomy. A typical CCD has high quantum efficiency, which enables it to detect a large percentage of the photons which impinge upon its surface. However, a CCD is generally not suitable for imaging under extremely low light conditions. This is mainly due to the measurement noise of the CCD, which can overwhelm the signal when relatively few photons are detected from the imaged object. Measurement noise is added when the signal is read out from the device, and is commonly called the readout noise (e.g., [Snyder et al. 1995](#)).

A technology that is intended to overcome the detrimental effect of the readout noise under low light conditions is the electron-multiplying charge-coupled device (EMCCD). Similar to a CCD, an EMCCD is a device which accumulates electrons in proportion to the number of photons it detects. Unlike a CCD, however, an EMCCD can significantly increase the number of electrons through a multiplication process before they are read out. Therefore, whereas the readout noise can overwhelm the signal detected by a CCD under low light conditions, it is effectively drowned out by the amplified signal in the case of an EMCCD.

An EMCCD achieves signal amplification by transferring the electrons through a multiplication register that consists of a sequence of typically several hundred stages. Electrons are moved through the stages in sequence, repeatedly exiting one stage and entering the next until they exit the last stage. Multiplication occurs because in any given stage, each input electron can, with certain probabilities, generate secondary electrons via a physical phenomenon called impact ionization. All secondary electrons are transferred to the next stage for further amplification along with the input electrons. In this way, even when the probabilities of secondary electron generation (typically 0.01 to 0.02 ([Basden et al. 2003](#)) for one secondary electron per input electron, and even smaller for more than one per input electron) are small in each stage, the cascading effect can potentially produce a large number of electrons per initial electron. An important observation is that while multiplication is intended to drown out the readout noise, it is a random process, and hence introduces stochasticity of its own. Therefore, when enough photons can be detected from the imaged object such that the readout noise level is already low in comparison to the signal, the use of electron multiplication can actually have the adverse effect of unnecessarily introducing increased stochasticity in the acquired data.

From an image acquired using a CCD-based camera, parameters of interest can be estimated to obtain useful information about the imaged object. In single molecule microscopy (e.g., [Walter et al. 2008](#); [Moerner 2007](#)), for example, a major topic of interest has been the accurate estimation of the location of a fluorescent molecule (e.g., [Thompson et al. 2002](#); [Ober et al. 2004](#); [Andersson 2008](#); [Pavani and Piestun 2008](#); [Ram et al. 2008](#)), which has important applications in, for instance, the study of intracellular processes (e.g., [Ram et al. 2008](#)). In this context, we have proposed a general framework ([Ram et al. 2006](#)) for calculating the Fisher information, and hence the Cramer-Rao lower bound ([Rao 1965](#)), for the estimation of parameters from an image produced by a microscope. Based on this framework, accuracy limits have been obtained for estimating, for example, the positional coordinates of a single molecule (e.g., [Ober et al. 2004](#)). These performance measures, however, assume the image to have been acquired by a CCD, and hence do not apply to EMCCD imaging.

Our current purpose is then to develop the theory that is necessary for deriving performance measures for estimating parameters from an EMCCD image. To obtain the Fisher information for EMCCD data, we need to find the probability distribution of the electron count

that results from the multiplication process described above. To this end, we consider the stochastic multiplication as a branching process (e.g., [Harris 1963](#); [Athreya and Ney 2004](#)), as others have done for EMCCDs (e.g., [Hynecek and Nishiwaki 2003](#); [Basden et al. 2003](#); [Tubbs 2003](#)) and other electron-multiplying devices (e.g., [Matsuo et al. 1985](#); [Hollenhorst 1990](#)). The probability distribution for the output of the branching process depends on the model that is used to describe the secondary electron generation in each multiplication stage. While multiple secondary electrons can be generated per input electron per stage ([Hynecek and Nishiwaki 2003](#)), for simplicity it is common to describe the electron generation as a Bernoulli event (i.e., a secondary electron is generated per input electron per stage with probability  $b$ , or not generated with probability  $1 - b$ ). However, since no explicit expression exists for the probability distribution of the output electron count of such a branching process, different approximating distributions have been proposed (e.g., [Hynecek and Nishiwaki 2003](#); [Basden et al. 2003](#); [Tubbs 2003](#)). Rather than starting with the assumption of Bernoulli multiplication, here we propose using the geometric distribution to describe the probability of secondary electron generation.

The main contribution of the current work lies in the derivation of Fisher information expressions for data that can be described as the output of a branching process, with or without added readout noise. The expressions differ in the random variable, and hence the probability distribution, that is used to model the data. In all cases, however, the input to the branching process is modeled as a Poisson random variable. This assumption of a Poisson-distributed initial signal is significant, in that many physical processes that give rise to stochastic signals can be modeled as a Poisson process. Photon emission by a fluorescent molecule (and hence the detection of those photons by a camera), for example, is typically modeled as a Poisson process.

Importantly, in the process of presenting our theory, illustrations and comparisons are made of different data models in terms of their Fisher information. To facilitate the investigation, we use a quantity which we call the “noise coefficient”. The noise coefficient allows us to analyze and compare the Fisher information of different data models via a scalar quantity, and its definition is based on the assumption that information about the parameters of interest is contained in the Poisson-distributed signal.

We note that in the context of EMCCD image analysis, little has been done, to the best of our knowledge, in the way of Fisher information which importantly allows the computation of Cramer-Rao lower bound-based accuracy limits for estimating parameters of interest from the acquired data. The studies performed in [Hynecek and Nishiwaki \(2003\)](#) and [Basden et al. \(2003\)](#), for example, consider the stochasticity of electron multiplication, but not from the perspective of Fisher information. Note however that in [Mortensen et al. \(2010\)](#), the Fisher information has been presented for an EMCCD data model that is characterized by a probability distribution that has previously been described in [Basden et al. \(2003\)](#). This distribution is also the probability density for what we refer to as the high gain approximation (see Sect. 5.1) of our geometric model of electron multiplication. In their analysis, Mortensen et al. ultimately arrive at a simple approximation of the Fisher information for EMCCD data that is based on the “excess noise factor,” a well known quantity in electron multiplication literature (e.g., [Matsuo et al. 1985](#); [Hollenhorst 1990](#); [Hynecek and Nishiwaki 2003](#)). Specifically, the Fisher information for EMCCD data is approximated as half the Fisher information for the ideal scenario where the data consists of just the initial Poisson-distributed signal. However, as suggested in their work, as well as by our analysis here (see Sect. 5.2), this approximation is appropriate only when a relatively large number of photons are detected by an EMCCD operating at a high level of amplification. In contrast, the Fisher information corresponding to our

geometric model of multiplication is suitable for arbitrary photon counts and amplification levels.

This paper is organized as follows. We begin in Sect. 2 with a general result from which all ensuing Fisher information expressions can be derived. This result immediately gives rise to the Fisher information for the ideal scenario of an uncorrupted signal and the practical scenario of a signal that is corrupted by readout noise. Based on this result, we also define the noise coefficient. In Sect. 3, general Fisher information expressions are presented which pertain to a signal that is amplified according to a branching process. These results assume a general model of multiplication, and allow the straightforward derivation of expressions for specific multiplication models. In Sect. 4, explicit Fisher information expressions are derived for the geometric model of multiplication. Using these results, we examine multiplication in terms of the stochasticity that it introduces and its ability to drown out the device readout noise. In Sect. 5, Fisher information expressions are presented for two approximations of an amplified signal that are relatively easy to work with computationally and/or analytically. In Sect. 6, we generalize the theory developed in previous sections for a single signal to a collection of independent signals such as that which comprise an EMCCD image. This is followed by an example which applies the generalized theory to the localization of a single molecule from an EMCCD image. In this example, the calculation of the Fisher information matrix and the Cramer-Rao lower bound-based limit of the localization accuracy is accompanied by the results of maximum likelihood estimations performed with simulated images. Finally, we conclude in Sect. 7.

## 2 The noise coefficient

The first part of this paper deals with the analysis of the information content of a scalar random variable that models the data in a single pixel of an electron multiplication imaging sensor or a similar device. The signal that impacts the detector is modeled as a Poisson random variable, but the data is a readout noise-corrupted version of the signal that may have been amplified with the intention to overcome the added noise.

Of particular interest in this paper is the calculation of the Fisher information matrix related to various parameter estimation problems such as the localization of a single molecule from its image. These estimation problems take on a typical form. The probability mass function of the incident Poisson random variable is parameterized by its mean  $\nu$ . The mean  $\nu$  itself is, however, typically a function of the parameter vector  $\theta$  that is of interest, e.g., the positional coordinates of a single molecule. Since these problems form the basis of what follows, we first state an expression for the Fisher information of a random variable using the specific parameterization that will be of importance for the later developments.

**Theorem 1** *Let  $Z_\nu$  be a continuous (discrete) random variable with probability density (mass) function  $p_\nu$ , where  $\nu$  is a scalar parameter. Let  $\nu = \nu_\theta$  be a reparameterization of  $p_\nu$  through the possibly vector-valued parameter  $\theta \in \Theta$ , where  $\Theta$  is the parameter space. We use the notation  $Z_\theta(p_\theta)$  to denote  $Z_{\nu_\theta}(p_{\nu_\theta})$ ,  $\theta \in \Theta$ . Then the Fisher information matrix  $\mathbf{I}(\theta)$  of  $Z_\theta$  with respect to  $\theta$  is given by*

$$\mathbf{I}(\theta) = \left( \frac{\partial \nu_\theta}{\partial \theta} \right)^T \frac{\partial \nu_\theta}{\partial \theta} \cdot E \left[ \left( \frac{\partial}{\partial \nu_\theta} \ln(p_\theta(z)) \right)^2 \right]. \quad (1)$$

*Proof* Let  $L(\theta|z)$  be the log-likelihood function for  $\theta$ . By the definition of the Fisher information matrix,

$$\begin{aligned} \mathbf{I}(\theta) &= E \left[ \left( \frac{\partial L(\theta|z)}{\partial \theta} \right)^T \frac{\partial L(\theta|z)}{\partial \theta} \right] = E \left[ \left( \frac{\partial}{\partial \theta} \ln(p_\theta(z)) \right)^T \frac{\partial}{\partial \theta} \ln(p_\theta(z)) \right] \\ &= \left( \frac{\partial v_\theta}{\partial \theta} \right)^T \frac{\partial v_\theta}{\partial \theta} \cdot E \left[ \left( \frac{1}{p_\theta(z)} \frac{\partial p_\theta(z)}{\partial v_\theta} \right)^2 \right] = \left( \frac{\partial v_\theta}{\partial \theta} \right)^T \frac{\partial v_\theta}{\partial \theta} \cdot E \left[ \left( \frac{\partial}{\partial v_\theta} \ln(p_\theta(z)) \right)^2 \right]. \end{aligned}$$

□

Note that the above theorem can be seen to be a special case of the transformation result for Fisher information matrices (e.g., Kay 1993) that applies to our specific conditions, and that the scalar expectation term in Eq. 1 is just the Fisher information of  $Z_\theta$  with respect to  $v_\theta$ .

Theorem 1 importantly breaks the Fisher information matrix  $\mathbf{I}(\theta)$  into two parts: a matrix portion  $\left( \frac{\partial v_\theta}{\partial \theta} \right)^T \frac{\partial v_\theta}{\partial \theta}$  which does not depend on the specific probability distribution  $p_\theta$  of  $Z_\theta$ , and a scalar portion  $E \left[ \left( \frac{\partial}{\partial v_\theta} \ln(p_\theta(z)) \right)^2 \right]$  which does depend on  $p_\theta$ . This formulation provides the foundation of our definition of a “noise coefficient” for comparing the Fisher information of different data models. To motivate the purpose and definition of this coefficient, we first present two corollaries which follow immediately from Theorem 1. The results of both corollaries have been presented in our earlier work (e.g., Ober et al. 2004; Ram et al. 2006), but are given here again as two important realizations of the general Fisher information expression of Eq. 1.

The two corollaries pertain to data acquired under two important scenarios. The first gives the Fisher information matrix for the ideal scenario where a Poisson-distributed number of detected particles are read out from a device without being corrupted by measurement noise. This scenario thus represents the best case wherein a CCD is able to read out a signal without introducing readout noise, and will serve as the benchmark against which practical scenarios are compared. In this and all ensuing scenarios, the function  $v_\theta$  will represent the mean of the Poisson-distributed signal.

**Corollary 1** *Let  $Z_\theta$  be a Poisson random variable with mean  $v_\theta > 0$ , such that its probability mass function is  $p_{\theta,P}(z) = \frac{e^{-v_\theta} v_\theta^z}{z!}$ ,  $z = 0, 1, \dots$ . The Fisher information matrix  $\mathbf{I}_P(\theta)$  of  $Z_\theta$  is given by*

$$\mathbf{I}_P(\theta) = \left( \frac{\partial v_\theta}{\partial \theta} \right)^T \frac{\partial v_\theta}{\partial \theta} \cdot \frac{1}{v_\theta}. \tag{2}$$

*Proof* Since  $p_{\theta,P}$  is parameterized by  $\theta$  through the function  $v_\theta$ , by Theorem 1  $\mathbf{I}_P(\theta)$  takes the form of Eq. 1. To prove the result, we need only evaluate the expectation term of Eq. 1, with  $p_\theta = p_{\theta,P}$ :

$$\begin{aligned} E \left[ \left( \frac{\partial}{\partial v_\theta} [\ln(p_{\theta,P}(z))] \right)^2 \right] &= E \left[ \left( \frac{\partial}{\partial v_\theta} \left[ \ln \left( \frac{e^{-v_\theta} v_\theta^z}{z!} \right) \right] \right)^2 \right] = E \left[ \left( \frac{z}{v_\theta} - 1 \right)^2 \right] \\ &= \frac{1}{v_\theta^2} E[z^2] - \frac{2}{v_\theta} E[z] + 1 = \frac{v_\theta + v_\theta^2}{v_\theta^2} - 2 \frac{v_\theta}{v_\theta} + 1 = \frac{1}{v_\theta}. \end{aligned}$$

□

Note that Eq. 2 reduces to the familiar result of  $\mathbf{I}_P(\theta) = \frac{1}{v_\theta}$  if the parameter to be estimated happens to be the mean of the Poisson distribution (i.e., if  $\theta = v_\theta$ ).

The second corollary is of significance in that it gives the Fisher information matrix for the practical scenario where measurement noise is added to a Poisson-distributed number of detected particles when they are read out from a device. Measurement noise is typically modeled as a Gaussian random variable, and is assumed to be such by this corollary (see Appendix for the proof). This scenario importantly corresponds to the practical operation of a CCD, which adds measurement noise when electrons are read out (e.g., Snyder et al. 1995).

**Corollary 2** Let  $Z_\theta = V_\theta + W$ , where  $V_\theta$  is a Poisson random variable with mean  $\nu_\theta > 0$ , and  $W$  is a Gaussian random variable with mean  $\eta_w$  and variance  $\sigma_w^2$ . Let  $V_\theta$  and  $W$  be stochastically independent of each other, and let  $W$  be not dependent on  $\theta$ . The probability density function of  $Z_\theta$  is then the convolution of the Poisson probability mass function with mean  $\nu_\theta$  and the Gaussian probability density function with mean  $\eta_w$  and variance  $\sigma_w^2$ , given by  $p_{\theta,R}(z) = \frac{1}{\sqrt{2\pi}\sigma_w} \sum_{j=0}^\infty \frac{e^{-\nu_\theta} \nu_\theta^j}{j!} e^{-\frac{1}{2}\left(\frac{z-j-\eta_w}{\sigma_w}\right)^2}$ ,  $z \in \mathbb{R}$ . The Fisher information matrix  $\mathbf{I}_R(\theta)$  of  $Z_\theta$  is given by

$$\mathbf{I}_R(\theta) = \left(\frac{\partial v_\theta}{\partial \theta}\right)^T \frac{\partial v_\theta}{\partial \theta} \cdot \left(\int_{-\infty}^\infty \frac{1}{p_{\theta,R}(z)} \cdot \left(\frac{1}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \frac{e^{-\nu_\theta} \nu_\theta^{j-1}}{(j-1)!} e^{-\frac{1}{2}\left(\frac{z-j-\eta_w}{\sigma_w}\right)^2}\right)^2 dz - 1\right). \tag{3}$$

Note that the density function  $p_{\theta,R}$  in the above corollary can be found in Snyder et al. (1995).

According to Theorem 1 and demonstrated by Corollaries 1 and 2, different probability distributions of the random variable  $Z_\theta$  will result in Fisher information matrices that differ from one another through the Fisher information of  $Z_\theta$  with respect to  $\nu_\theta$  (i.e., the scalar expectation term of Eq. 1). We therefore introduce, for the purpose of comparing the Fisher information of different data models, a “noise coefficient” based on this quantity. (We call it the “noise coefficient” because, as illustrated by Corollaries 1 and 2, different data models will typically share the same signal, but differ by the type of noise that corrupts the signal.) Since Eq. 2 represents the best case scenario (where we have Poisson-distributed data that is not corrupted by measurement noise), we take its Fisher information with respect to  $\nu_\theta$  (i.e., the quantity  $\frac{1}{\nu_\theta}$ ) as the reference, and define the noise coefficient as follows.

**Definition 1** Let  $Z_\theta$  be a continuous (discrete) random variable with probability density (mass) function  $p_\theta$ . Let  $p_\theta$  be parameterized by  $\theta$  through the mean  $\nu_\theta > 0$  of a Poisson-distributed random variable. Then the *noise coefficient* (with respect to  $\nu_\theta$ ) of  $Z_\theta$ , denoted by  $\alpha$ , is given by

$$\alpha = \nu_\theta \cdot E \left[ \left( \frac{\partial}{\partial \nu_\theta} \ln(p_\theta(z)) \right)^2 \right]. \tag{4}$$

The noise coefficient of Eq. 4 is just the ratio of the Fisher information of  $Z_\theta$  to that of the ideal, uncorrupted Poisson-distributed signal, both with respect to  $\nu_\theta$ . Using this quantity, the Fisher information matrix  $\mathbf{I}(\theta)$  of a random variable which satisfies the conditions of Definition 1 can be expressed as  $\mathbf{I}(\theta) = \alpha \cdot \mathbf{I}_P(\theta)$ , where  $\mathbf{I}_P(\theta)$  is the Fisher information matrix of Eq. 2. Note that the noise coefficient  $\alpha$  as defined is a nonnegative scalar, and that the larger its value, the larger the Fisher information matrix  $\mathbf{I}(\theta)$  (i.e., the higher the amount of information  $\mathbf{I}(\theta)$  contains about the parameter vector  $\theta$ , and hence the better the accuracy with which  $\theta$  can be estimated).

For the ideal scenario of Corollary 1 where the data is Poisson-distributed, the noise coefficient is trivially  $\alpha_P = 1$ . For the practical scenario of Corollary 2 where Gaussian-distributed measurement noise is added to the detected Poisson-distributed signal, the noise coefficient  $\alpha_R$  is just

$$\alpha_R = \nu_\theta \cdot \left( \int_{-\infty}^{\infty} \frac{1}{p_{\theta,R}(z)} \cdot \left( \frac{1}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^{\infty} \frac{e^{-\nu_\theta} \nu_\theta^{j-1}}{(j-1)!} e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} \right)^2 dz - 1 \right), \tag{5}$$

and the Fisher information matrix  $\mathbf{I}_R(\theta)$  can be written as  $\mathbf{I}_R(\theta) = \alpha_R \cdot \mathbf{I}_P(\theta)$ .

In this paper, most results will involve data that can be described as a Poisson-distributed signal with mean  $\nu_\theta$  that is stochastically multiplied (i.e., amplified) by some random function  $M$  before being corrupted by some additive measurement noise  $W$ . Since neither the stochasticity introduced by the multiplication nor the measurement noise is dependent on  $\theta$ , they contribute no additional information about  $\theta$ . If anything, they can deteriorate the Fisher information. Therefore, the noise coefficient  $\alpha$  for these data models can be expected to be at most 1 (i.e., at most  $\alpha_P$ ). We state and prove this result formally in the following theorem.

**Theorem 2** *Let  $\Theta$  be a parameter space and let  $Z_\theta = M(V_\theta) + W, \theta \in \Theta$ , where  $V_\theta$  is a Poisson random variable with mean  $\nu_\theta > 0$ ,  $M$  is a random function, and  $W$  is a scalar-valued random variable. We assume that  $V_\theta, M$ , and  $W$  are stochastically independent. Then, for the noise coefficient  $\alpha$  of  $Z_\theta, 0 \leq \alpha \leq 1$ .*

*Proof* Trivially, we have  $\alpha \geq 0$  because by Definition 1, it is the product of a positive value  $\nu_\theta$  and the expectation of a squared expression. We next show that  $\alpha \leq 1$ .

For any random variable  $U_\theta$  that depends on the parameter  $\theta \in \Theta$ , we denote by  $S_{U_\theta}$  its score function and by  $\mathbf{I}_{U_\theta}(\theta) = E \left[ (S_{U_\theta})^T S_{U_\theta} \right]$  its Fisher information matrix with respect to the parameter  $\theta$ . If we let  $Y_\theta = M(V_\theta)$ , we first show that  $\mathbf{I}_{Y_\theta}(\theta) \leq \mathbf{I}_{V_\theta}(\theta)$ .

We explicitly denote the dependence of the multiplication random function  $M$  on the random event  $\omega \in \Omega$ , i.e.,  $M : \mathbb{N} \times \Omega \rightarrow \mathbb{R}; (n, \omega) \mapsto M(n, \omega)$ . For a fixed  $\omega_0 \in \Omega, M(\cdot, \omega_0)$  can be considered as a constant statistic for the Poisson random variable  $V_\theta$ . Then by the monotonicity of the Fisher information matrix (e.g., Rao 1965), we have that for each  $\omega_0 \in \Omega, \mathbf{I}_{M(V_\theta, \omega_0)}(\theta) \leq \mathbf{I}_{V_\theta}(\theta)$ .

Using the identity  $E[E[A|B]] = E[A]$ , where  $A$  and  $B$  are random variables, we condition on the random events that yield constant (i.e., non-random) multiplication functions to obtain the required inequality

$$\begin{aligned} \mathbf{I}_{Y_\theta}(\theta) &= \mathbf{I}_{M(V_\theta)}(\theta) = E \left[ (S_{M(V_\theta)})^T S_{M(V_\theta)} \right] = E \left[ E \left[ (S_{M(V_\theta)})^T S_{M(V_\theta)} | M \right] \right] \\ &= E \left[ E \left[ (S_{M(V_\theta)})^T S_{M(V_\theta)} | M(\cdot, \cdot) = M(\cdot, \omega_0) \right] \right] \\ &= E \left[ \mathbf{I}_{M(V_\theta, \omega_0)}(\theta) \right] \leq E \left[ \mathbf{I}_{V_\theta}(\theta) \right] = \mathbf{I}_{V_\theta}(\theta). \end{aligned}$$

We next consider the addition of the random variable  $W$  to  $Y_\theta$ . Let  $F$  be the statistic defined by  $F([A \ B]^T) = [1 \ 1][A \ B]^T = A + B$ , where  $A$  and  $B$  are random variables. We then have that  $Z_\theta = F([Y_\theta \ W]^T)$ , and by the monotonicity of the Fisher information matrix, we have  $\mathbf{I}_{Z_\theta}(\theta) \leq \mathbf{I}_{[Y_\theta \ W]^T}(\theta)$ .

By the stochastic independence of  $Y_\theta$  and  $W$ , the joint probability density function of  $[Y_\theta \ W]^T$  is the product of their respective density functions  $p_{Y_\theta}$  and  $p_W$ . (Note that  $Y_\theta$  and  $W$  can each be either discrete or continuous, but for ease of treatment, we consider them both

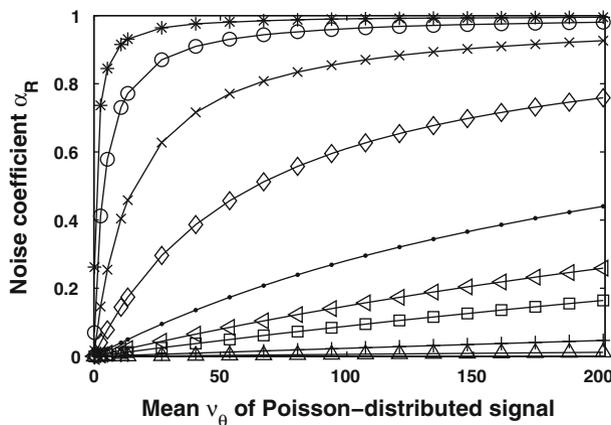
to be characterized by density functions.) Then,

$$\begin{aligned} \mathbf{I}_{[Y_\theta \ w]^T}(\theta) &= E \left[ \left( \frac{\partial}{\partial \theta} \ln[p_{Y_\theta}(y)p_W(w)] \right)^T \frac{\partial}{\partial \theta} \ln[p_{Y_\theta}(y)p_W(w)] \right] \\ &= E \left[ \left( \frac{\partial}{\partial \theta} \ln[p_{Y_\theta}(y)] + \frac{\partial}{\partial \theta} \ln[p_W(w)] \right)^T \left( \frac{\partial}{\partial \theta} \ln[p_{Y_\theta}(y)] + \frac{\partial}{\partial \theta} \ln[p_W(w)] \right) \right] \\ &= E \left[ \left( \frac{\partial}{\partial \theta} \ln[p_{Y_\theta}(y)] \right)^T \left( \frac{\partial}{\partial \theta} \ln[p_{Y_\theta}(y)] \right) \right] = \mathbf{I}_{Y_\theta}(\theta). \end{aligned}$$

Therefore, we have  $\mathbf{I}_{Z_\theta}(\theta) \leq \mathbf{I}_{Y_\theta}(\theta) \leq \mathbf{I}_{V_\theta}(\theta)$ . Since  $\mathbf{I}_{V_\theta}(\theta)$  is just  $\mathbf{I}_P(\theta)$  of Corollary 1, and since  $\mathbf{I}_{Z_\theta}(\theta) = \alpha \cdot \mathbf{I}_P(\theta)$ , it follows that  $\alpha \leq 1$ .  $\square$

Note that if the multiplication random function  $M$  in Theorem 2 is the identity function (i.e.,  $M(V_\theta) = V_\theta$ ), then we recover the signal plus measurement noise scenario of Corollary 2. Hence, Theorem 2 immediately implies that the noise coefficient  $\alpha_R$  of Eq. 5 is between 0 and 1, and that the Fisher information matrix  $\mathbf{I}_R(\theta)$  is no greater than  $\mathbf{I}_P(\theta)$  of the ideal scenario of Corollary 1. For illustration, Fig. 1 plots, for  $\eta_w = 0$  and different values of  $\sigma_w$  corresponding to different levels of readout noise, the noise coefficient  $\alpha_R$  as a function of the mean  $\nu_\theta$  of the Poisson-distributed signal. The plot shows that for all readout noise levels and for all values of  $\nu_\theta$ , the value of the noise coefficient  $\alpha_R$  is no less than 0 and no greater than 1.

For each value of  $\nu_\theta$ , Fig. 1 shows that the value of  $\alpha_R$  decreases with increasing levels of the readout noise. This agrees with the expectation that the Fisher information deteriorates in proportion to the amount of noise. Moreover, for each readout noise level, we see that the value of  $\alpha_R$  increases with increasing values of  $\nu_\theta$ . This demonstrates that, by increasing the signal to effectively drown out the readout noise, the Fisher information matrix  $\mathbf{I}_R(\theta)$  can be expected to approach  $\mathbf{I}_P(\theta)$  of the ideal scenario. Practically, however, it may be impossible to acquire the necessary amount of signal. In this case, the signal can be amplified before it



**Fig. 1** Noise coefficient  $\alpha_R$  (Eq. 5), for the scenario of a Poisson-distributed signal with Gaussian-distributed measurement noise, as a function of the mean  $\nu_\theta$  of the signal. The values of  $\nu_\theta$  range from 0.27 to 201.28. The different curves correspond to measurement noise with mean  $\eta_w = 0$  and standard deviation  $\sigma_w = 1(*), 2(\circ), 4(\times), 8(\diamond), 16(\cdot), 24(\triangleleft), 32(\square), 64(+),$  and  $128(\triangle)$

is read out. Signal amplification that can be modeled as a branching process is the subject of what follows.

### 3 Fisher information for signal multiplication

In a basic multiplication, or branching, process (e.g., [Harris 1963](#); [Athreya and Ney 2004](#)), an initial number of particles are fed into a series of stages where, in each stage, an input particle can generate secondary particles of the same kind with certain probabilities. The particles at the output of one stage become the input of the next, resulting in a cascading effect that can produce a large number of particles, and hence an amplification of the initial particle count. While motivated by electron multiplication in an EMCCD, the theory we present is general and can be used for other multiplication applications. Therefore, consistent with the literature on branching processes, we will use the generic term “particle” to refer to the physical entities that are multiplied. We begin by giving the definition of a basic branching process, in a form similar to that found in [Kimmel and Axelrod \(2002\)](#).

**Definition 2** A branching process with an initial particle count probability distribution  $(q_k)_{k=0,1,\dots}$  and an individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$  is a sequence of nonnegative integer-valued random variables  $(X_n)_{n=0,1,\dots}$ , given by

$$X_n = \begin{cases} U, & n = 0, \\ \sum_{j=1}^{X_{n-1}} Y_j, & n = 1, 2, \dots, \end{cases} \tag{6}$$

where the initial particle count random variable  $U$  is nonnegative and integer-valued with probability distribution  $(q_k)_{k=0,1,\dots}$ , and for each  $X_n, n = 1, 2, \dots$ , the individual offspring count random variables  $Y_j, j = 1, \dots, X_{n-1}$ , are nonnegative and integer-valued, and are mutually independent and identically distributed with probability distribution  $(p_k)_{k=0,1,\dots}$ . Note that if  $X_{n-1} = 0$ , then  $\sum_{j=1}^{X_{n-1}} Y_j = 0$ .

In [Definition 2](#), the random variable  $X_n$  denotes the number of particles at the output of the  $n$ th stage of multiplication, with  $X_0$  understood to be the initial number of particles  $U$  prior to multiplication. In the definition of  $X_n$  for  $n = 1, 2, \dots$ , the random variable  $Y_j, j = 1, \dots, X_{n-1}$ , denotes the number of offspring particles due to the  $j$ th of the  $X_{n-1}$  particles at the output (input) of the previous (current) stage. In this way, the cascading effect of the multiplication process is captured. Note that the number of offspring particles  $Y_j$  includes the  $j$ th input particle itself.

For a branching process with  $N$  stages of multiplication, where  $N$  is a nonnegative integer, the mean and variance of the number of particles  $X_N$  at the output of the  $N$ th stage are well known results (e.g., [Harris 1963](#); [Athreya and Ney 2004](#)). For a general initial particle count probability distribution  $(q_k)_{k=0,1,\dots}$  (which by [Definition 2](#) has mean  $E[X_0]$  and variance  $Var[X_0]$ ) and a general individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$  with mean  $m$  and variance  $\sigma^2$ , the mean  $E[X_N]$  and variance  $Var[X_N]$  of  $X_N$  are as given in [Table 1](#). The table also gives expressions for  $E[X_N]$  and  $Var[X_N]$  that correspond to combinations of different definitions of  $(q_k)_{k=0,1,\dots}$  and  $(p_k)_{k=0,1,\dots}$ . These expressions are easily verified by substituting  $E[X_0] = Var[X_0] = v_\theta$ , and/or appropriate expressions for  $m$  and  $\sigma^2$  (see [Sect. 4](#) for definitions of the standard and zero modified geometric distributions), into the general expressions for  $E[X_N]$  and  $Var[X_N]$ .

From [Table 1](#), we see that  $E[X_N]$  is the product of the mean initial particle count  $E[X_0]$  and the term  $m^N$ , which is the mean of  $X_N$  given a single initial particle (i.e., given that

**Table 1** The mean  $E[X_N]$  and variance  $Var[X_N]$  of the number of particles  $X_N$  at the output of an  $N$ -stage branching process for various combinations of the initial particle count probability distribution  $(q_k)_{k=0,1,\dots}$  and the individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$

$(q_k)_{k=0,1,\dots}$	$(p_k)_{k=0,1,\dots}$	$E[X_N]$	$Var[X_N]$
General	General	$m^N E[X_0]$	$\frac{\sigma^2 m^{N-1} (m^N - 1)}{m - 1} E[X_0] + m^{2N} Var[X_0]$
Poisson( $v_\theta$ )	General	$m^N v_\theta$	$v_\theta \left( \frac{\sigma^2 m^{N-1} (m^N - 1)}{m - 1} + m^{2N} \right)$
Poisson( $v_\theta$ )	Geometric( $b$ )	$\left( \frac{1}{1-b} \right)^N v_\theta$	$v_\theta \left( 2 \left( \frac{1}{1-b} \right)^{2N} - \left( \frac{1}{1-b} \right)^N \right)$
Poisson( $v_\theta$ )	Zero modified geometric( $a, b$ )	$\left( \frac{1-a}{1-b} \right)^N v_\theta$	$v_\theta \left( \frac{2b}{b-d} \left( \frac{1-a}{1-b} \right)^{2N} - \frac{b+d}{b-d} \left( \frac{1-a}{1-b} \right)^N \right)$

The mean and variance of a general  $(p_k)_{k=0,1,\dots}$  are denoted by  $m$  and  $\sigma^2$ , where  $m \geq 0$  for  $E[X_N]$ , and  $m > 0, m \neq 1$  for  $Var[X_N]$

$X_0 = 1$ ). The term  $m^N$  is commonly called the *mean gain* of the branching process, and is an indicator of the amount of multiplication that can be expected from a branching process.

Definition 2 specifies a general branching process by leaving the initial particle count probability distribution  $(q_k)_{k=0,1,\dots}$  and the individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$  to be defined. We are, however, interested in the scenario where the initial particle count random variable is Poisson-distributed (e.g., photons detected as electrons by an EMCCD according to a Poisson process). Using the Poisson distribution for  $(q_k)_{k=0,1,\dots}$ , the next theorem gives the probability distribution, the noise coefficient, and the Fisher information matrix of the particle count  $X_{N,\theta}$  at the output of an  $N$ -stage branching process.

**Theorem 3** Let  $X_{N,\theta}, N \in \{0, 1, \dots\}$ , be the number of particles at the output of an  $N$ -stage branching process with an initial particle count probability distribution  $q_k = \frac{e^{-v_\theta} v_\theta^k}{k!}, k = 0, 1, \dots$  and an individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$  that is not dependent on  $\theta$ .

1. The probability distribution of  $X_{N,\theta}$  is given by the probability mass function

$$p_{\theta,M}(x) = P(X_{N,\theta} = x) = \sum_{j=0}^{\infty} P(X_{N,\theta} = x | X_{0,\theta} = j) \cdot \frac{e^{-v_\theta} v_\theta^j}{j!}, \quad x = 0, 1, \dots \quad (7)$$

2. The noise coefficient corresponding to the probability mass function  $p_{\theta,M}$  is given by

$$\alpha_M = v_\theta \cdot \left( \sum_{x=0}^{\infty} \frac{1}{P(X_{N,\theta} = x)} \left( \sum_{j=1}^{\infty} P(X_{N,\theta} = x | X_{0,\theta} = j) \cdot \frac{e^{-v_\theta} v_\theta^{j-1}}{(j-1)!} \right)^2 - 1 \right), \quad (8)$$

and the Fisher information matrix of  $X_{N,\theta}$  is  $\mathbf{I}_M(\theta) = \alpha_M \cdot \mathbf{I}_P(\theta)$ , where  $\mathbf{I}_P(\theta)$  is as given in Eq. 2.

*Proof* 1. By the law of total probability, for  $x = 0, 1, \dots$ ,

$$\begin{aligned}
 p_{\theta, M}(x) &= P(X_{N, \theta} = x) = \sum_{j=0}^{\infty} P(X_{N, \theta} = x | X_{0, \theta} = j) \cdot q_j \\
 &= \sum_{j=0}^{\infty} P(X_{N, \theta} = x | X_{0, \theta} = j) \cdot \frac{e^{-\nu_{\theta}} \nu_{\theta}^j}{j!}.
 \end{aligned}$$

2. Since the individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$  is not dependent on  $\theta$ , the conditional probability  $P(X_{N, \theta} = x | X_{0, \theta} = j)$ ,  $j = 0, 1, \dots$ , is not dependent on  $\theta$ . Hence,  $p_{\theta, M}$  is parameterized by  $\theta$  through the mean  $\nu_{\theta}$  of a Poisson-distributed random variable. By Definition 1,  $\alpha_M$  is then given by Eq. 4 with  $p_{\theta} = p_{\theta, M}$ , and the resulting expectation term can be evaluated analogously in step-by-step fashion as in the proof of Corollary 2. The matrix  $\mathbf{I}_M(\theta)$  then follows directly from Definition 1.  $\square$

Equations 7 and 8 are expressed in terms of  $P(X_{N, \theta} = x | X_{0, \theta} = j)$ , which is the general probability mass function of the output particle count  $X_{N, \theta}$  conditioned on the initial particle count  $X_{0, \theta}$ . They therefore allow the straightforward realization of concrete expressions for the probability mass function and the noise coefficient of  $X_{N, \theta}$  once the conditional mass function is determined. The conditional mass function, in turn, needs to be defined through a specific individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$ , which specifies how the particles are multiplied. Indeed, this approach is taken in Sect. 4 to arrive at the Fisher information matrix for the case where  $(p_k)_{k=0,1,\dots}$  is a geometric probability distribution.

Theorem 3 assumes that the output particle count  $X_{N, \theta}$  of the multiplication process is the data from which  $\theta$  is to be estimated. In a device such as an EMCCD, however, measurement noise is added to  $X_{N, \theta}$  during the readout process. In this case, the data  $Z_{\theta}$  is the sum of  $X_{N, \theta}$  and a Gaussian random variable  $W$  which models the noise. By the stochastic independence of  $X_{N, \theta}$  and  $W$ , the density function  $p_{\theta, MR}$  (Eq. 31) of  $Z_{\theta}$  is the convolution of the mass function of  $X_{N, \theta}$  (Eq. 7) and a Gaussian density function. Using  $p_{\theta, MR}$ , application of Definition 1 and a calculation analogous to the proof of Corollary 2 will readily yield the noise coefficient  $\alpha_{MR}$  (Eq. 32). These results are stated as Corollary 4 in the Appendix.

### 4 Fisher information for geometric multiplication

Theorem 3 and Corollary 4 provide general Fisher information expressions for a Poisson-distributed initial particle count that has been stochastically multiplied according to some individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$ . In this section, we consider the case where  $(p_k)_{k=0,1,\dots}$  is a zero modified geometric distribution, which we define as follows (Haccou et al. 2005).

**Definition 3** A zero modified geometric distribution is a probability distribution  $(p_k)_{k=0,1,\dots}$  given by

$$p_k = \begin{cases} a, & k = 0, \\ (1 - a)(1 - b)b^{k-1}, & k = 1, 2, \dots, \end{cases} \quad 0 \leq a < 1, \quad 0 \leq b < 1. \tag{9}$$

The zero modified geometric distribution has mean  $m = \frac{1-a}{1-b}$  and variance  $\sigma^2 = \frac{(1-a)(b+a)}{(1-b)^2}$ , which can be easily verified. It has a probability generating function of linear fractional form. Due to this special property, the probability distribution of the number of particles  $X_{N, \theta}$  at

the output of an  $N$ -stage branching process with a geometrically distributed individual offspring count can be expressed explicitly without recursion (e.g., Harris 1963; Athreya and Ney 2004). From Eq. 9, it can be seen that the zero modified geometric distribution allows the individual offspring count to take on increasing positive integer values with decreasing probabilities, and if desired, the value zero with a non-zero probability. This makes it suitable for modeling processes like electron multiplication in an EMCCD, where higher offspring counts per input particle per stage can be assumed to become increasingly unlikely. (Note that this assumption is inferred from Hyneczek and Nishiwaki (2003), where it is suggested that the generation of more than one secondary electron per input electron per stage in an EMCCD multiplication register can be assumed to be an event that practically does not occur.) The ability to define a non-zero probability for a zero individual offspring count also enables the modeling of any particle loss mechanisms that may exist.

By setting the parameter  $a = 0$ , the zero modified geometric distribution of Eq. 9 reduces to the standard geometric distribution  $p_k = (1 - b)b^{k-1}$ ,  $k = 1, 2, \dots$ . This distribution has easily verifiable mean  $m = \frac{1}{1-b}$  and variance  $\sigma^2 = \frac{b}{(1-b)^2}$ . We will use the more general zero modified geometric distribution in the theory that follows. However, the standard geometric distribution will be used for illustration purposes since it produces a branching process where there is no possibility of losing a particle. The absence of particle loss is often assumed in the modeling of, for example, electron multiplication in an EMCCD.

For an  $N$ -stage branching process with a Poisson-distributed initial particle count and a geometrically distributed individual offspring count, the following theorem gives the probability distribution, the noise coefficient, and the Fisher information matrix of the output particle count  $X_{N,\theta}$ . In brief, the proof (see Appendix) for this theorem uses the theory of probability generating functions to derive the conditional probability mass function of  $X_{N,\theta}$  given an initial particle count  $X_{0,\theta}$  (see Eq. 34). The results then follow by using the conditional mass function in the general mass function and noise coefficient of Eqs. 7 and 8.

**Theorem 4** *Let  $X_{N,\theta}$ ,  $N \in \{0, 1, \dots\}$ , be the number of particles at the output of an  $N$ -stage branching process with an initial particle count probability distribution  $q_k = \frac{e^{-v_\theta} v_\theta^k}{k!}$ ,  $k = 0, 1, \dots$ , and a zero modified geometric individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$  that is not dependent on  $\theta$ .*

1. *The probability distribution of  $X_{N,\theta}$  is given by the probability mass function*

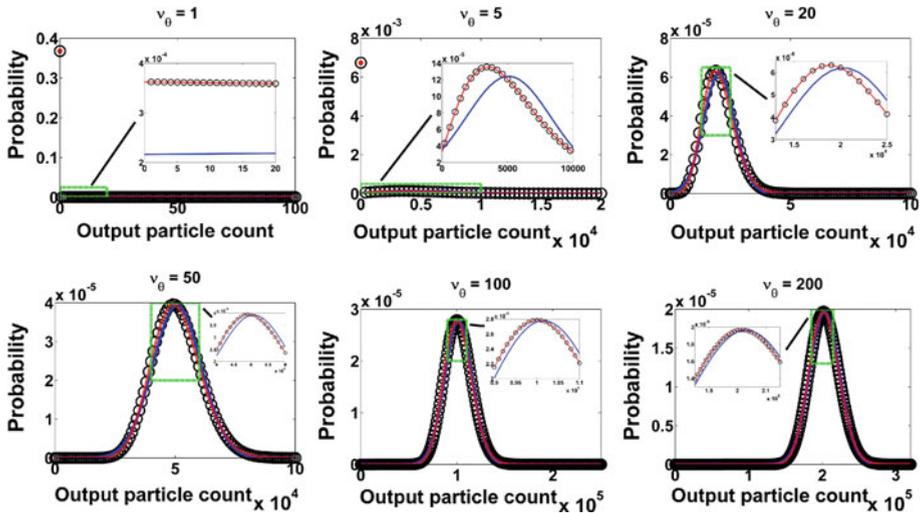
$$p_{\theta,Geom}(x) = P(X_{N,\theta} = x) = \begin{cases} e^{-v_\theta} \frac{A}{B}, & x = 0, \\ e^{-v_\theta} \frac{A}{B} \sum_{j=0}^{x-1} \binom{x-1}{j} \frac{C^{x-1-j} D^{j+1}}{B^{j+x+1}} \frac{v_\theta^{j+1}}{(j+1)!}, & x = 1, 2, \dots, \end{cases} \tag{10}$$

where  $A = (1 - a)(m - 1)m^N$ ,  $B = b(m^N - 1)m + (1 - a)(m - 1)$ ,  $C = b(m^N - 1)m$ ,  $D = m^N(1 - a)^2(m - 1)^2$ ,  $m = \frac{1-a}{1-b} \neq 1$ , and  $\binom{x-1}{j}$  denotes “ $x - 1$  choose  $j$ ”.

2. *The noise coefficient corresponding to the probability mass function  $p_{\theta,Geom}$  is given by*

$$\alpha_{Geom} = v_\theta \cdot \left[ \sum_{x=1}^{\infty} \frac{e^{-2v_\theta} \frac{A}{B}}{p_{\theta,Geom}(x)} \left( \sum_{j=0}^{x-1} \binom{x-1}{j} \frac{C^{x-1-j} D^{j+1}}{B^{j+x+1}} \frac{v_\theta^j}{j!} \right)^2 - \left( \frac{A}{B} \right)^2 \right], \tag{11}$$

and the Fisher information matrix of  $X_{N,\theta}$  is  $\mathbf{I}_{Geom}(\theta) = \alpha_{Geom} \cdot \mathbf{I}_P(\theta)$ , with  $\mathbf{I}_P(\theta)$  as given in Eq. 2.



**Fig. 2** Probability mass function (Eq. 12) of  $X_{536,\theta}$ , the particle count at the output of a 536-stage branching process with a standard geometric ( $\circ$ ) individual offspring count probability distribution and a mean gain of  $m^{536} = 1015.46$ , for six different values of the mean  $v_\theta$  of the Poisson-distributed initial particle count. For each value of  $v_\theta$ , the probability density functions resulting from the high gain approximation (Eq. 21) (red curve) and the Gaussian approximation (see discussion of Theorem 6) (blue curve) of the output of the standard geometric multiplication are also shown

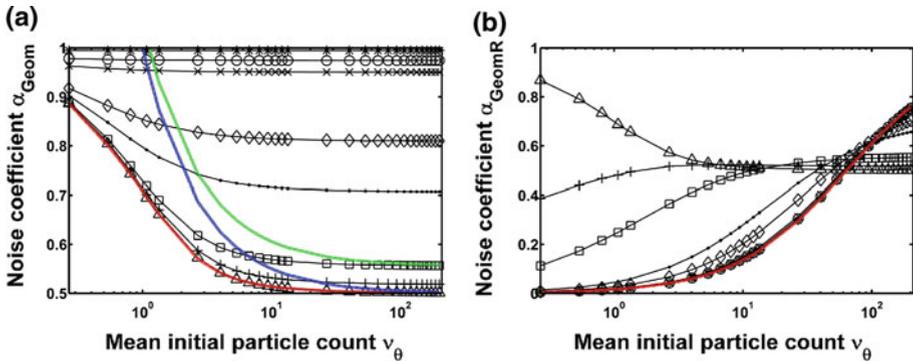
For the special case where we have a standard geometric individual offspring count probability distribution, setting  $a = 0$  in Eqs. 10 and 11 yields, for  $m = \frac{1}{1-b}$ , the simpler expressions

$$p_{\theta,Geom}(x) = \begin{cases} e^{-v_\theta}, & x = 0, \\ e^{-v_\theta} \sum_{j=0}^{x-1} \frac{\binom{x-1}{j}}{(j+1)!} \left(1 - \frac{1}{m^N}\right)^{x-1-j} \left(\frac{v_\theta}{m^N}\right)^{j+1}, & x = 1, 2, \dots, \end{cases} \quad (12)$$

$$\alpha_{Geom} = v_\theta \cdot \left[ \sum_{x=1}^{\infty} \frac{e^{-2v_\theta}}{p_{\theta,Geom}(x)} \left( \sum_{j=0}^{x-1} \binom{x-1}{j} \left(1 - \frac{1}{m^N}\right)^{x-1-j} \left(\frac{1}{m^N}\right)^{j+1} \frac{v_\theta^j}{j!} \right)^2 - 1 \right]. \quad (13)$$

Figure 2 plots, in open circles, the probability mass function  $p_{\theta,Geom}(x)$  of Eq. 12 for six different values of the mean initial particle count  $v_\theta$ . For small values of  $v_\theta$  (1 and 5), the shape of the function is characterized by a large value at  $x = 0$ . This is because the output particle count of the multiplication will be 0 whenever the initial particle count is 0, a likely event given the small  $v_\theta$ . However, as  $v_\theta$  becomes larger, the shape of the probability mass function starts to resemble that of a Gaussian density function (see Sect. 5.2).

Multiplication is a random process, and hence adds stochasticity to the detected signal. This is illustrated by Fig. 3a, which plots the noise coefficient  $\alpha_{Geom}$  of Eq. 13 as a function of the mean  $v_\theta$  of the initial particle count. The figure shows curves for different values of the mean gain  $m^N$  of the multiplication process, with  $N$  set to 536 stages. (The value 536 is the number of multiplication stages in a CCD97 chip (E2V Technologies, Chelmsford, UK), a sensor which is used in many EMCCD cameras.) In accordance with Theorem 2, the stochasticity added by the multiplication is seen in the fact that the noise coefficient  $\alpha_{Geom}$  is less than one for all values of the mean gain and  $v_\theta$  shown in the plot. This directly implies



**Fig. 3** **a** Noise coefficient  $\alpha_{Geom}$  (Eq. 13), for the scenario of a Poisson-distributed signal that is amplified by multiplication, and **b** noise coefficient  $\alpha_{GeomR}$  (Eq. 38), for the scenario of a Poisson-distributed signal that is amplified by multiplication and subsequently corrupted by measurement noise. In both **a** and **b**, the noise coefficient is shown as a function of the mean  $\nu_\theta$  of the signal, which ranges in value from 0.27 to 201.28. The signal is amplified through  $N = 536$  stages of standard geometric multiplication, and the *different curves* correspond to mean gain values of  $m^N = 1.01(*), 1.03(\circ), 1.06(\times), 1.31(\diamond), 1.71(\cdot), 4.98(\square), 14.49(+),$  and  $1015.46(\Delta)$ . In **a**, the *red curve* represents the noise coefficient  $\alpha_H$  (Eq. 22) of the high gain approximation, while the *blue and green curves* represent the noise coefficient  $\alpha_{Gauss}$  (Eq. 25) of the Gaussian approximation for mean gains of 1015.46 and 4.98, respectively. In **b**, the measurement noise is Gaussian-distributed with mean  $\eta_w = 0$  and standard deviation  $\sigma_w = 8$ , and as a reference, the *red curve* represents the noise coefficient  $\alpha_R$  (Eq. 5) for the scenario where there is no signal amplification

that the corresponding Fisher information matrix  $\mathbf{I}_{Geom}(\theta)$  is smaller than  $\mathbf{I}_P(\theta)$  of the ideal scenario.

For each value of  $\nu_\theta$ , Fig. 3a shows that multiplication with a higher mean gain results in a smaller value of  $\alpha_{Geom}$ , indicating that the amount of added stochasticity is proportional to the mean gain. Moreover, for each mean gain value, we see that the value of  $\alpha_{Geom}$  decreases with increasing values of  $\nu_\theta$ , suggesting that multiplication adds the least amounts of stochasticity when the signal to be amplified is small.

Despite the fact that multiplication adds stochasticity to the data, it can be useful in the practical scenario wherein measurement noise is introduced by the device. In this case, the derivations of the probability density  $p_{\theta,GeomR}$  (Eq. 35) of the final data  $Z_\theta$  and the associated noise coefficient  $\alpha_{GeomR}$  (Eq. 36) will be analogous to the proof of Theorem 4 in the Appendix, using instead the general expressions of Eqs. 31 and 32 of Corollary 4, which account for the added Gaussian noise. The results are stated as Corollary 5 in the Appendix, where expressions for the special case of standard geometric multiplication are also given.

To illustrate that multiplication can be beneficial when readout noise is present, Fig. 3b plots the noise coefficient  $\alpha_{GeomR}$  of Eq. 38 as a function of the mean  $\nu_\theta$  of the initial particle count. Curves are shown for the same set of mean gain values as in Fig. 3a, and a zero mean Gaussian readout noise with a standard deviation of  $\sigma_w = 8$  is assumed. As a reference for comparison, the noise coefficient  $\alpha_R$  of Eq. 5, which corresponds to the absence of multiplication, is shown for the same readout noise level as a red curve.

Figure 3b shows that, for the given set of mean gain values, multiplication yields a better noise coefficient than no multiplication (i.e.,  $\alpha_{GeomR}$  is greater than  $\alpha_R$ ) for  $\nu_\theta$  values of up to roughly 60 particles. In this range of  $\nu_\theta$  values, a higher mean gain generally yields a larger  $\alpha_{GeomR}$ . Beyond roughly  $\nu_\theta = 60$  particles, however, multiplication starts to produce, in order of decreasing mean gain, a noise coefficient that is worse than no multiplication (i.e.,  $\alpha_{GeomR}$  starts to drop below  $\alpha_R$ ). By roughly  $\nu_\theta = 130$  particles, multiplication with

any of the given mean gain values offers no benefit, with a higher mean gain yielding a smaller  $\alpha_{GeomR}$ .

Figure 3b therefore demonstrates that multiplication, especially at high mean gain, is beneficial when the signal level is relatively small (or equivalently, when the readout noise level is relatively significant). On the other hand, when the signal level is relatively large such that the readout noise level is already relatively insignificant, multiplication has the undesirable net effect of introducing additional stochasticity.

### 5 Approximations of the output of multiplication

Though exact, the probability distributions and noise coefficients of Theorem 4 and Corollary 5 can be difficult to analyze and time-consuming to compute. In this section, we present two ways to approximate the output of geometric multiplication that yield expressions that are relatively simple to analyze and/or compute. The first approximation is based on the behavior of a branching process as its mean gain converges to infinity, and applies specifically to geometric multiplication. The second approximation models the output of a branching process as a Gaussian random variable, and can be used with any multiplication model. This approximation allows for a relatively easy analytical demonstration of the behavior of the noise coefficient.

#### 5.1 High gain approximation

Let  $(X_{n,\theta})_{n=0,1,\dots}$  be a branching process with initial particle count  $X_{0,\theta} = 1$  and a zero modified geometric individual offspring count probability distribution with mean  $m = \frac{1-a}{1-b} > 1$ . Then it is shown in Harris (1963) that as the number of multiplication stages  $n$  converges to infinity, the sequence of random variables  $(Y_n = X_{n,\theta}/m^n)_{n=0,1,\dots}$  converges, with probability 1 and in mean square, to a random variable  $Y$  with a zero modified exponential distribution given by

$$p(y) = \begin{cases} \frac{a}{b}, & y = 0, \\ \left(1 - \frac{a}{b}\right)^2 e^{-(1-\frac{a}{b})y}, & y > 0, \end{cases} \quad 0 \leq a < b < 1. \tag{14}$$

Based on this result, for a large  $N$ , the particle count  $X_{N,\theta}$  at the output of an  $N$ -stage branching process with initial particle count  $X_{0,\theta} = 1$  and a zero modified geometric individual offspring count probability distribution with mean  $m > 1$ , can be approximated by the random variable  $m^N Y$ , whose probability density function can be verified to be a scaled version of Eq. 14:

$$p(x) = \begin{cases} \frac{a}{b}, & x = 0, \\ \frac{1}{m^N} \left(1 - \frac{a}{b}\right)^2 e^{-(1-\frac{a}{b})\frac{x}{m^N}}, & x > 0, \end{cases} \quad 0 \leq a < b < 1. \tag{15}$$

If instead the branching process has an initial particle count of  $X_{0,\theta} = j, j = 1, 2, \dots$ , then the distribution of the output particle count  $X_{N,\theta}$  can be approximated by the  $j$ -fold convolution of  $p(x)$  of Eq. 15 with itself. The resulting density function can be shown to be, for  $0 \leq a < b < 1$ ,

$$p(x|j) = \begin{cases} \left(\frac{a}{b}\right)^j, & x = 0, \\ \frac{1}{m^{Nj}} \left(1 - \frac{a}{b}\right)^2 e^{-(1-\frac{a}{b})\frac{x}{m^{Nj}}} \sum_{l=0}^{j-1} \frac{\binom{j-1}{l} j}{(l+1)!} \left(\frac{a}{b}\right)^{j-l-1} \left(1 - \frac{a}{b}\right)^{2l} \left(\frac{x}{m^{Nj}}\right)^l, & x > 0. \end{cases} \tag{16}$$

To obtain the approximating probability density function and noise coefficient for the case where the initial particle count  $X_{0,\theta}$  is Poisson-distributed with parameter  $\nu_\theta$ , we note that Eq. 16 is the approximating probability density function of the output particle count  $X_{N,\theta}$  conditioned on  $X_{0,\theta}$ . It is thus similar to the conditional probability mass function that is required by the general expressions of Eqs. 7 and 8. However, since it is a density function, continuous analogues of Eqs. 7 and 8 are needed. By proceeding in similar fashion as in the proof of Theorem 3, the general expressions for the continuous case can be verified to be

$$p(x) = \begin{cases} e^{-\nu_\theta} \left( 1 + \sum_{j=1}^\infty \frac{p(0|j)v_\theta^j}{j!} \right), & x = 0, \\ e^{-\nu_\theta} \sum_{j=1}^\infty \frac{p(x|j)v_\theta^j}{j!}, & x > 0, \end{cases} \tag{17}$$

$$\alpha = \int_0^\infty \frac{1}{p(x)} \left( e^{-\nu_\theta} \sum_{j=1}^\infty \frac{p(x|j)v_\theta^{j-1}}{(j-1)!} \right)^2 dx + \frac{1}{p(0)} \left( e^{-\nu_\theta} \sum_{j=1}^\infty \frac{p(0|j)v_\theta^{j-1}}{(j-1)!} \right)^2 - 1. \tag{18}$$

Substituting Eq. 16 into Eq. 17 gives the required density function, which we use in our next theorem.

**Theorem 5** *Let  $X_\theta$  be a random variable with probability density function*

$$p_{\theta,H}(x) = \begin{cases} e^{-\nu_\theta(1-\frac{a}{b})}, & x = 0, \\ \frac{(1-\frac{a}{b})^2 e^{-(1-\frac{a}{b})\frac{x}{m^N} - \nu_\theta}}{m^N} \sum_{j=1}^\infty \frac{v_\theta^j}{(j-1)!} \sum_{k=0}^{j-1} \frac{\binom{j-1}{k} (\frac{a}{b})^{j-k-1}}{(k+1)! \left( (1-\frac{a}{b})^2 \frac{x}{m^N} \right)^{-k}}, & x > 0, \end{cases} \tag{19}$$

where  $0 \leq a < b < 1$  and  $m = \frac{1-a}{1-b}$ . Then the noise coefficient corresponding to  $p_{\theta,H}$  is

$$\alpha_H = \nu_\theta \cdot \left[ \int_0^\infty \frac{\left( \sum_{j=1}^\infty \frac{j v_\theta^{j-1}}{(j-1)!} \sum_{k=0}^{j-1} \frac{\binom{j-1}{k} (\frac{a}{b})^{j-k-1}}{(k+1)! \left( (1-\frac{a}{b})^2 x \right)^{-k}} \right)^2}{e^{(1-\frac{a}{b})x + \nu_\theta} \sum_{j=1}^\infty \frac{v_\theta^j}{(j-1)!} \sum_{k=0}^{j-1} \frac{\binom{j-1}{k} (\frac{a}{b})^{j-k-1}}{(k+1)! \left( (1-\frac{a}{b})^2 x \right)^{-k}}} dx + \frac{a^2}{b^2 e^{\nu_\theta(1-\frac{a}{b})}} - 1 \right], \tag{20}$$

and the Fisher information matrix of  $X_\theta$  is  $\mathbf{I}_H(\theta) = \alpha_H \cdot \mathbf{I}_P(\theta)$ , where  $\mathbf{I}_P(\theta)$  is as given in Eq. 2.

*Proof* The noise coefficient  $\alpha_H$  follows by substituting Eqs. 19 and 16 into Eq. 18 and simplifying. The Fisher information matrix  $\mathbf{I}_H(\theta)$  then follows directly from Definition 1.  $\square$

To approximate standard geometric multiplication, setting  $a = 0$  in Eqs. 19 and 20 yields the simpler expressions

$$p_{\theta,H}(x) = \begin{cases} e^{-v_{\theta}}, & x = 0, \\ \frac{e^{-\left(\frac{x}{m^N} + v_{\theta}\right)} \sqrt{v_{\theta}x/m^N} I_1\left(2\sqrt{v_{\theta}x/m^N}\right)}{x}, & x > 0, \end{cases} \quad m = \frac{1}{1-b}, \quad 0 < b < 1, \tag{21}$$

$$\alpha_H = v_{\theta} \cdot \left[ \frac{e^{-v_{\theta}}}{4v_{\theta}^2} \int_0^{\infty} \frac{x^2 e^{-\frac{x^2}{4v_{\theta}}} I_0^2(x)}{I_1(x)} dx - 1 \right]. \tag{22}$$

The functions  $I_1$  and  $I_0$  are the first and zeroth order modified Bessel function of the first kind, respectively, and they are obtained by applying the identity (Abramowitz and Stegun 1965)

$$I_v(u) = \left(\frac{1}{2}u\right)^v \sum_{j=0}^{\infty} \frac{\left(\frac{1}{4}u^2\right)^j}{j!(v+j)!}, \quad v \in \{0, 1, \dots\}, \quad u \in \mathbb{R},$$

with  $v = 1$  and  $v = 0$  during the simplification. Note that Eq. 21 appears in an equivalent form, without the jump at  $x = 0$  and without the modified Bessel function representation, in Basden et al. (2003). It can also be found, in identical form as presented here and along with some associated Fisher information analysis, in Mortensen et al. (2010). Note also that whereas the mean gain  $m^N$  is present in the density functions of Eqs. 19 and 21, it disappears by simple substitution in deriving the noise coefficients of Eqs. 20 and 22.

According to its derivation, the high gain approximation can be expected to apply only in the case of a large mean gain. For illustration, the probability density function of Eq. 21 is shown, for a relatively large mean gain of  $m^N = 1015.46$ , as a red curve in each plot of Fig. 2. For each of the six values of the mean initial particle count  $v_{\theta}$ , a close agreement can be seen between this density function and the probability mass function of Eq. 12 (o). Similarly, a close match can be seen in Fig. 3a between the noise coefficient  $\alpha_H$  of Eq. 22 (red curve) and the noise coefficient  $\alpha_{Geom}$  of Eq. 13 computed for  $m^N = 1015.46$  ( $\Delta$ ). As expected, Fig. 3a also demonstrates that the agreement between  $\alpha_H$  (which again does not depend on the mean gain  $m^N$ ) and  $\alpha_{Geom}$  becomes increasingly poor with decreasing mean gain values for  $\alpha_{Geom}$ .

When measurement noise is present, the approximating probability distribution of the final data will be the convolution of the approximating distribution of Eq. 19 and a Gaussian distribution. The resulting density function  $p_{\theta,HR}$  (Eq. 39) and the associated noise coefficient  $\alpha_{HR}$  (Eq. 40) are stated as Corollary 6 in the Appendix, where expressions pertaining to standard geometric multiplication are also given.

The high gain approximation offers the important advantage of computational efficiency. At least for the case of standard geometric multiplication, we have found that its associated expressions can be evaluated in significantly less time than their exact counterparts in Sect. 4. Importantly, as demonstrated above, they produce values which are similar to that of their exact counterparts at high mean gain. In Sect. 6.2.2, we will make use of this approximation when performing maximum likelihood estimations where the probability distribution of the data is computed many times with different values of the estimated parameters.

### 5.2 Gaussian approximation

In this section, we present a Gaussian approximation of the output of multiplication. Unlike the high gain approximation, its applicability is not restricted to geometric multiplication. It has been reported (e.g., Hynecek and Nishiwaki 2003; Basden et al. 2003) that given a large initial particle count, the probability distribution of the output particle count  $Y$  of a branching process can be approximated by a Gaussian distribution with the same mean and variance as that of  $Y$ . For an  $N$ -stage branching process with a Poisson initial particle count probability distribution and a general individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$ , the mean and variance of the number of particles  $X_{N,\theta}$  at its output are given in Table 1. Taking this mean and variance to be that of a Gaussian random variable  $X_\theta$ , the following theorem gives the noise coefficient and the Fisher information matrix of the particle count  $Z_\theta = X_\theta + W$ , where  $W$  is a Gaussian random variable representing measurement noise. In contrast to previous sections, here we are presenting the more general case first because the case without measurement noise is a trivial special case of the more general result.

**Theorem 6** *Let  $Z_\theta = X_\theta + W$ , where  $X_\theta$  and  $W$  are Gaussian random variables that are stochastically independent of each other, and  $W$  is not dependent on  $\theta$ . Let  $X_\theta$  have mean  $\eta_{x,\theta} = m^N v_\theta$  and variance  $\sigma_{x,\theta}^2 = v_\theta \left( \frac{\sigma^2 m^{N-1} (m^N - 1)}{m - 1} + m^{2N} \right)$ , where  $v_\theta > 0, m > 0, m \neq 1, \sigma^2 \geq 0$ , and  $N \in \{0, 1, \dots\}$ . Let  $W$  have mean  $\eta_w$  and variance  $\sigma_w^2$ . Then the noise coefficient corresponding to the probability density function of  $Z_\theta$ , which is Gaussian with mean  $\eta_{z,\theta} = \eta_{x,\theta} + \eta_w$  and variance  $\sigma_{z,\theta}^2 = \sigma_{x,\theta}^2 + \sigma_w^2$ , is given by*

$$\alpha_{GaussR} = v_\theta \cdot \left[ \left( v_\theta \left( \frac{\sigma^2 \left( 1 - \frac{1}{m^N} \right)}{m(m-1)} + 1 \right) + \frac{\sigma_w^2}{m^{2N}} \right)^{-1} + \frac{1}{2} \left( v_\theta + \frac{\sigma_w^2 m^{-2N}}{\frac{\sigma^2 \left( 1 - \frac{1}{m^N} \right)}{m(m-1)} + 1} \right)^{-2} \right], \tag{23}$$

and the Fisher information matrix of  $Z_\theta$  is  $\mathbf{I}_{GaussR}(\theta) = \alpha_{GaussR} \cdot \mathbf{I}_P(\theta)$ , with  $\mathbf{I}_P(\theta)$  as given in Eq. 2.

*Proof* By Definition 1,  $\alpha_{GaussR}$  is the product of  $v_\theta$  and the Fisher information of the Gaussian random variable  $Z_\theta$  with respect to  $v_\theta$ . The Fisher information of a Gaussian random variable is well known (e.g., Kay 1993). For  $Z_\theta$  with mean  $\eta_{z,\theta}$  and variance  $\sigma_{z,\theta}^2$ , its Fisher information with respect to  $v_\theta$  is

$$\mathbf{I}(\theta) = \frac{1}{\sigma_{z,\theta}^2} \left( \frac{\partial \eta_{z,\theta}}{\partial v_\theta} \right)^2 + \frac{1}{2\sigma_{z,\theta}^4} \left( \frac{\partial \sigma_{z,\theta}^2}{\partial v_\theta} \right)^2. \tag{24}$$

By evaluating Eq. 24 and multiplying the result by  $v_\theta$ , we arrive at Eq. 23. □

For the case where measurement noise is absent, the random variable  $Z_\theta$  of Theorem 6 reduces to the Gaussian random variable  $X_\theta$  as defined in the theorem. By making this adjustment in the proof of Theorem 6, one can easily verify the resulting noise coefficient to be, for  $m > 0, m \neq 1$ , and  $\sigma^2 \geq 0$ ,

$$\alpha_{Gauss} = \left( \frac{\sigma^2 \left( 1 - \frac{1}{m^N} \right)}{m(m-1)} + 1 \right)^{-1} + \frac{1}{2v_\theta}. \tag{25}$$

The Fisher information matrix of  $X_\theta$  is then just  $\mathbf{I}_{Gauss}(\theta) = \alpha_{Gauss} \cdot \mathbf{I}_P(\theta)$ , with  $\mathbf{I}_P(\theta)$  as given in Eq. 2.

In electron multiplication literature (e.g., Matsuo et al. 1985; Hollenhorst 1990; Hynecek and Nishiwaki 2003), the *excess noise factor* is defined as the ratio of the variance of the output particle count of a multiplication process to the variance of the initial particle count, normalized by the square of the mean gain. It is thus a measure of the increase in the stochasticity of the data as a result of the multiplication. Interestingly, the expression  $\frac{\sigma^2(1-1/m^N)}{m(m-1)} + 1$  in Eqs. 23 and 25 is easily verifiable to be the excess noise factor for the case of a Poisson-distributed initial particle count. We will revisit the excess noise factor later in this section.

By substituting  $m = \frac{1}{1-b}$  and  $\sigma^2 = \frac{b}{(1-b)^2}$  for the standard geometric distribution into the mean  $\eta_{x,\theta}$  and the variance  $\sigma_{x,\theta}^2$  of  $X_\theta$ , the resulting Gaussian density function is plotted, for a mean gain of  $m^{536} = 1015.46$ , as a blue curve in Fig. 2 for each of the six values of the mean initial particle count  $\nu_\theta$ . We see that for small values of  $\nu_\theta$ , the Gaussian density matches poorly with the probability mass function of Eq. 12 (o). However, with increasing values of  $\nu_\theta$ , the Gaussian density starts to coincide more and more with the mass function. The same observation can be made with the noise coefficient  $\alpha_{Gauss}$  of Eq. 25. Computed for  $m = \frac{1}{1-b}$ ,  $\sigma^2 = \frac{b}{(1-b)^2}$ , and  $m^{536} = 1015.46$ , Fig. 3a shows that, for small values of  $\nu_\theta$ ,  $\alpha_{Gauss}$  (blue curve) is in poor agreement with  $\alpha_{Geom}$  of Eq. 13 that has been computed for the same mean gain ( $\Delta$ ). With increasing values of  $\nu_\theta$ , however, it becomes an increasingly better match of  $\alpha_{Geom}$ . Hence, unlike the high gain approximation which applies for all values of  $\nu_\theta$ , the Gaussian approximation only applies for large  $\nu_\theta$  as indicated in Hynecek and Nishiwaki (2003) and Basden et al. (2003). However, unlike the high gain approximation, the Gaussian approximation also applies for low mean gain values. For example, Fig. 3a shows that for a low mean gain of  $m^{536} = 4.98$ ,  $\alpha_{Gauss}$  (green curve) becomes an increasingly better match of  $\alpha_{Geom}$  ( $\square$ ) with increasing values of  $\nu_\theta$ .

As is evident from Eq. 25 and its curves in Fig. 3a, the Gaussian approximation can produce noise coefficients that converge to infinity when  $\nu_\theta$  converges to 0. This might appear to contradict Theorem 2, which states that the noise coefficient can be at most 1. However, this is not the case since the Gaussian random variable does not fall under the category of random variables to which the theorem applies. Specifically, it is not the sum of a random function of a Poisson random variable and a scalar-valued random variable.

Besides being easy to compute, the expressions associated with the Gaussian approximation allow for relatively simple analytical studies. For example, it is easily verified that as the mean gain  $m^N$  converges to infinity in Eqs. 23 and 25, in both cases we obtain the same noise coefficient

$$\alpha_{GaussR} = \alpha_{Gauss} = \left( \frac{\sigma^2}{m(m-1)} + 1 \right)^{-1} + \frac{1}{2\nu_\theta}, \quad m > 1. \tag{26}$$

This shows that, in practice, we can drown out the measurement noise by using multiplication with a large mean gain to amplify the signal. Furthermore, the dependence of the noise coefficient on the inverse of  $\nu_\theta$  indicates that the high mean gain multiplication is more beneficial when low signal levels are expected. This is consistent with what we observed from Fig. 3b.

Using mean  $m = \frac{1}{1-b}$  and variance  $\sigma^2 = \frac{b}{(1-b)^2}$  for standard geometric multiplication, Eq. 26 yields

$$\alpha_{GaussR} = \alpha_{Gauss} = \frac{1}{2} + \frac{1}{2\nu_\theta}. \tag{27}$$

The first summand in Eq. 27 is the inverse of the excess noise factor (see above) in the limit that the mean gain converges to infinity. Therefore, in this limiting scenario, standard geometric multiplication yields an excess noise factor of 2, indicating a two-fold increase in the stochasticity of the data due to the amplification. Interestingly, for the Bernoulli multiplication model (see Sect. 1) which is commonly used to describe electron multiplication in an EMCCD, the excess noise factor has been shown to also approach the value 2 as the mean gain converges to infinity (e.g., Matsuo et al. 1985; Robbins and Hadwen 2003). However, the convergence to 2 for Bernoulli multiplication is dependent on the probability of secondary particle generation tending to 0.

If we further let the mean initial particle count  $\nu_\theta$  converge to infinity in Eq. 27, we obtain  $\alpha_{GaussR} = \alpha_{Gauss} = 0.5$ . This importantly suggests that when standard geometric multiplication with a high mean gain is unnecessarily used to amplify an already large signal, the Fisher information matrices  $\mathbf{I}_{GaussR}(\theta)$  and  $\mathbf{I}_{Gauss}(\theta)$  will be no smaller than half of  $\mathbf{I}_P(\theta)$  of the ideal Poisson data scenario. This agrees with Figs. 3a and b, which suggest that  $\alpha_{Geom}$  and  $\alpha_{GeomR}$  converge to 0.5 as the mean gain and  $\nu_\theta$  are increased.

## 6 Application to EMCCD imaging

When imaging with a CCD or an EMCCD, photons from the imaged object arrive, and hence are detected, according to a Poisson process. In each pixel of the device, electrons accumulate in proportion to the detected photons, and in the case of an EMCCD, are subsequently multiplied using a register consisting of several hundred stages. The readout process then converts the electric charge to a digital count, during which time measurement noise is added. The data in each pixel of a CCD (EMCCD) can therefore be modeled as a Poisson-distributed particle count (that is multiplied according to a branching process), plus a Gaussian-distributed random variable that models the readout noise. As such, the Fisher information expressions derived thus far will readily apply for a single CCD or EMCCD pixel. In this section, we first build on our theory by deriving results pertaining to the Fisher information for a collection of pixels (i.e., an image). We then give an example that applies the results to a concrete estimation problem in single molecule microscopy.

### 6.1 Fisher information for an EMCCD image

An image acquired by a CCD or an EMCCD consists of a two-dimensional array of pixels. Since the data in different pixels can be assumed to be independent measurements, the Fisher information matrix for an image is the sum of the Fisher information matrices for its pixels. For an image of  $K$  pixels, its Fisher information matrix is thus  $\mathbf{I}_{im}(\theta) = \sum_{k=1}^K \mathbf{I}_k(\theta)$ , where  $\mathbf{I}_k(\theta)$  is the Fisher information matrix for the  $k$ th pixel.

Moreover, since the theory from previous sections applies to the data in each pixel, we can write  $\mathbf{I}_{im}(\theta) = \sum_{k=1}^K \alpha_k \cdot \mathbf{I}_{P,k}(\theta)$ , where  $\alpha_k$  is the noise coefficient for the  $k$ th pixel, and  $\mathbf{I}_{P,k}(\theta) = \left( \frac{\partial \nu_{\theta,k}}{\partial \theta} \right)^T \frac{\partial \nu_{\theta,k}}{\partial \theta} \cdot \frac{1}{\nu_{\theta,k}}$  is the Fisher information matrix for the  $k$ th pixel in the ideal scenario where the data at that pixel is just the initial electron count which is Poisson-distributed with mean  $\nu_{\theta,k}$ . It follows that for the ideal  $K$ -pixel image,  $\alpha_k = 1$  for  $k = 1, \dots, K$ , and its Fisher information matrix is  $\mathbf{I}_{im,P}(\theta) = \sum_{k=1}^K \mathbf{I}_{P,k}(\theta)$ . We now give an inequality which relates  $\mathbf{I}_{im}(\theta)$  for a practical image to  $\mathbf{I}_{im,P}(\theta)$  for its corresponding ideal image.

**Theorem 7** Let  $\mathbf{I}_{im}(\theta) = \sum_{k=1}^K \alpha_k \cdot \mathbf{I}_{P,k}(\theta)$ , and let  $\mathbf{I}_{im,P}(\theta) = \sum_{k=1}^K \mathbf{I}_{P,k}(\theta)$ . Let  $\alpha_{min}$  and  $\alpha_{max}$  denote, respectively, the smallest and the largest elements in the sequence  $(\alpha_k)_{k=1,\dots,K}$ . Then we have

$$\alpha_{min} \cdot \mathbf{I}_{im,P}(\theta) \leq \mathbf{I}_{im}(\theta) \leq \alpha_{max} \cdot \mathbf{I}_{im,P}(\theta).$$

*Proof* Let  $\mathbf{J} = \mathbf{I}_{im}(\theta) - \alpha_{min} \cdot \mathbf{I}_{im,P}(\theta)$ . Then for the left-hand side inequality, we need to show that  $\mathbf{J} \geq 0$ . Writing out the matrices  $\mathbf{I}_{im}(\theta)$  and  $\mathbf{I}_{im,P}(\theta)$  explicitly as defined, we get

$$\mathbf{J} = \sum_{k=1}^K \alpha_k \cdot \mathbf{I}_{P,k}(\theta) - \alpha_{min} \sum_{k=1}^K \mathbf{I}_{P,k}(\theta) = \sum_{k=1}^K (\alpha_k - \alpha_{min}) \cdot \mathbf{I}_{P,k}(\theta).$$

Then, for any  $x \in \mathbb{R}^n$ , where  $\mathbf{J}$  is  $n \times n$ , we have that

$$x^T \mathbf{J} x = \sum_{k=1}^K x^T [(\alpha_k - \alpha_{min}) \cdot \mathbf{I}_{P,k}(\theta)] x = \sum_{k=1}^K (\alpha_k - \alpha_{min}) \cdot x^T \mathbf{I}_{P,k}(\theta) x \geq 0.$$

Note that we have the inequality at the end since, for all  $k$ ,  $(\alpha_k - \alpha_{min}) \geq 0$  and  $x^T \mathbf{I}_{P,k}(\theta) x \geq 0$ . The latter is true because for each  $k$ ,  $\mathbf{I}_{P,k}(\theta)$  is a Fisher information matrix, and therefore positive semidefinite. Since  $x^T \mathbf{J} x \geq 0$  for all  $x \in \mathbb{R}^n$ , we have that  $\mathbf{J} \geq 0$ . The proof for the right-hand side inequality is analogous.  $\square$

Theorem 7 is of importance in that it can be used to assess, in terms of the Fisher information, how close a practical image is to its corresponding ideal image. If, for example, the smallest noise coefficient for a practical image is close to 1, then the image contains nearly the same amount of information about  $\theta$  as the ideal image. On the other hand, if the largest noise coefficient is small compared to 1, then the image contains very little information about  $\theta$  compared to the ideal image. In general, provided that suitable models are used to compute  $\mathbf{I}_{im,P}(\theta)$  and the noise coefficients  $(\alpha_k)_{k=1,\dots,K}$ , Theorem 7 can be used to obtain a quantitative lower and upper bound of an image’s information content.

We now consider the case where the data in each pixel of a  $K$ -pixel EMCCD image is modeled using the Gaussian approximation of Theorem 6. Specifically, we assume the approximation of a standard geometric multiplication process with infinite mean gain, such that the noise coefficient for each pixel is in the form of  $\alpha_{GaussR}$  of Eq. 27. For this special scenario, we give a corollary which follows directly from Theorem 7.

**Corollary 3** Let  $\mathbf{I}_{im,GaussR}(\theta) = \sum_{k=1}^K \alpha_{GaussR,k} \cdot \mathbf{I}_{P,k}(\theta)$ , where for  $k = 1, \dots, K$ ,  $\alpha_{GaussR,k} = \frac{1}{2} + \frac{1}{2v_{\theta,k}}$ . Let  $\mathbf{I}_{im,P}(\theta) = \sum_{k=1}^K \mathbf{I}_{P,k}(\theta)$ , and let  $v_{\theta,max}$  and  $v_{\theta,min}$  denote, respectively, the smallest and the largest elements in the sequence  $(v_{\theta,k})_{k=1,\dots,K}$ . Then we have the inequality

$$\left(\frac{1}{2} + \frac{1}{2v_{\theta,max}}\right) \mathbf{I}_{im,P}(\theta) \leq \mathbf{I}_{im,GaussR}(\theta) \leq \left(\frac{1}{2} + \frac{1}{2v_{\theta,min}}\right) \mathbf{I}_{im,P}(\theta). \tag{28}$$

Corollary 3 is a realization of Theorem 7 which, due to the simpler form of the noise coefficient for the assumed conditions, gives an inequality that depends on just the maximum and the minimum mean initial electron counts in the pixels comprising the image. Importantly, Eq. 28 shows that if we let  $v_{\theta,k}$  converge to infinity for all pixels  $k = 1, \dots, K$ , then  $\mathbf{I}_{im,GaussR}(\theta) = \frac{1}{2} \mathbf{I}_{im,P}(\theta)$ . Practically, assuming that the data can be modeled using the Gaussian approximation, this indicates that when standard geometric multiplication is used at high mean gain to amplify an image with an already high signal level in every pixel, the Fisher information of the amplified image will at worst be half of that of its corresponding ideal image.

## 6.2 Example: limit of the accuracy for localizing a point source

In this section, we apply our theory to an important estimation problem in single molecule microscopy, namely the localization of a fluorescent molecule. We first compute and compare the noise coefficients for three different data models of an image acquired of a single molecule. These data models consist of the ideal scenario of an image with Poisson-distributed data, the practical scenario of a readout noise-corrupted CCD image, and the practical scenario of a stochastically multiplied and readout noise-corrupted EMCCD image. We then use the noise coefficients to compute Fisher information matrices for the estimation of the location of the single molecule from its image. From the matrices, we readily obtain the Cramer-Rao lower bound-based limits of the accuracy for the location estimation, which we compare with the standard deviations of maximum likelihood location estimates obtained with simulated image data sets. (For other studies involving point source parameter estimation from EMCCD images, see, for example, [DeSantis et al. \(2010\)](#), [Mortensen et al. \(2010\)](#), [Quan et al. \(2010\)](#), [Thompson et al. \(2010\)](#), and [Wu et al. \(2010\)](#).)

Note that in the material that follows, a standard geometric model of electron multiplication is assumed for the EMCCD. To perform the same analyses using a different model of multiplication, the results presented in Sect. 3 for a general model of multiplication may be used to obtain the necessary mathematical expressions.

### 6.2.1 Noise coefficient comparison

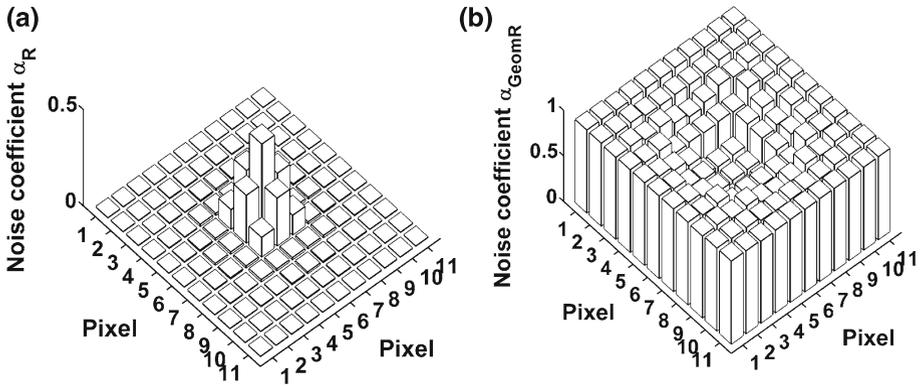
We consider the estimation of the location of an in-focus point source (e.g., single molecule) from its image as observed through a fluorescence microscope and detected by a CCD or EMCCD camera. For this problem, the mean of the Poisson-distributed photon count (or equivalently, electron count) that is detected from the point source at the  $k$ th pixel of the device can be shown to be

$$v_{\theta,k} = \frac{N_{\text{photon}}}{M^2} \int_{C_k} q(x/M - x_0, y/M - y_0) dx dy, \quad (29)$$

where  $N_{\text{photon}}$  is the expected number of photons detected from the point source,  $M$  is the magnification of the microscope,  $C_k$  is the region in the  $xy$ -plane occupied by the pixel,  $x_0$  and  $y_0$  are the  $x$  and  $y$  coordinates of the point source in the object space where it resides, and  $q$  is a function which describes the image formed from the detected photons. More specifically,  $q$  is called the “image function”, and gives the image at unit magnification of a point source that is located at the origin of the object space coordinate system ([Ram et al. 2006](#)). Equation 29 is a realization of a more general formula in [Ram et al. \(2006\)](#), as it assumes the rate at which photons are detected from the point source to be a constant. In general,  $v_{\theta,k}$  can also include as additional summands the means of Poisson-distributed components (e.g., photon count detected from something other than the point source) which are stochastically independent of one another, and of the electron count due to the point source.

We assume that  $q$  is given by the Airy point spread function ([Born and Wolf 1999](#)), a model which has classically been used to describe the image of an in-focus point source. The Airy image function can be written as ([Ram et al. 2006](#))

$$q(x, y) = \frac{J_1^2\left(\frac{2\pi n_a}{\lambda} \sqrt{x^2 + y^2}\right)}{\pi (x^2 + y^2)}, \quad (x, y) \in \mathbb{R}^2, \quad (30)$$



**Fig. 4** Noise coefficient ( $\alpha_R$  for **a**,  $\alpha_{GeomR}$  for **b**) profile for **a** a CCD image and **b** an EMCCD image of an in-focus point source. The point source is assumed to emit photons of wavelength  $\lambda = 655$  nm, which are collected by an objective lens with magnification  $M = 100$  and numerical aperture  $n_a = 1.3$ . The image of the point source is given by the Airy image function, and is centered on an 11-by-11 array of  $16 \mu\text{m}$  by  $16 \mu\text{m}$  pixels (i.e.,  $x_0 = y_0 = 880$  nm, assuming the upper left corner of the pixel array is  $(0, 0)$ ). The expected number of detected photons is set to  $N_{\text{photon}} = 200$ , a reasonable number that could be detected from a single fluorescent molecule. In **a**, readout noise with mean  $\eta_w = 0 e^-$  and standard deviation  $\sigma_w = 8 e^-$  is assumed for every pixel. In **b**, the standard deviation is higher at  $\sigma_w = 24 e^-$ , and standard geometric multiplication with a mean gain of  $m^{536} = 1015.46$  is assumed

where  $n_a$  is the numerical aperture of the objective lens,  $\lambda$  is the wavelength of the detected photons, and  $J_1$  is the first order Bessel function of the first kind. The Airy image function is circularly symmetric, with a strong central peak and ripples that increasingly diminish in amplitude away from the center.

Noise coefficients are computed for an 11-by-11 pixel array (i.e., image), with the mean initial electron count  $v_{\theta,k}$  (Eq. 29) at the  $k$ th pixel calculated with the point source attributes and imaging parameters given in Fig. 4. Mirroring the shape of the Airy image function of Eq. 30, and due to the centering of the point source image on the pixel array, the resulting profile of  $v_{\theta}$  values over the pixel array is circularly symmetric, with the largest values in the center region and a maximum value of 53.48 electrons at the center pixel.

For the ideal Poisson data model, the noise coefficient is, by definition, 1 for each pixel of the image. For the two practical data models, the noise coefficients for all pixels in the image are plotted in Fig. 4. For the CCD scenario, a Gaussian readout noise with mean  $\eta_w = 0 e^-$  and standard deviation  $\sigma_w = 8 e^-$  in every pixel is assumed. The chosen standard deviation represents a noise level that is typical of a CCD, and is the number specified, for example, for the normal readout mode of our laboratory’s Hamamatsu C4742-95-12ER CCD camera. In this case, Fig. 4a shows that  $\alpha_R$  of Eq. 5 is relatively small for every pixel. The center region of the image has the largest  $\alpha_R$  values, with the center pixel having the maximum value of only 0.456. This and the circular symmetry of the  $\alpha_R$  values correspond to the profile of  $v_{\theta}$  values, which is in accordance with Fig. 1 showing the value of  $\alpha_R$  to be proportional to the value of  $v_{\theta}$ . By Theorem 7, the maximum  $\alpha_R$  value of 0.456 implies that the Fisher information for this scenario is less than half of that for the ideal scenario.

For the EMCCD scenario, standard geometric multiplication using a 536-stage register is assumed. The mean gain is set to  $m^{536} = 1015.46$ , and a Gaussian readout noise with mean  $\eta_w = 0 e^-$  and standard deviation  $\sigma_w = 24 e^-$  in every pixel is assumed. The higher readout noise level of  $24 e^-$  is within the range of typical values for an EMCCD, and is representative

of the noise level for the settings at which our laboratory’s Andor iXon EMCCD cameras are often operated. In this case, Fig. 4b shows that the value of  $\alpha_{GeomR}$  of Eq. 38 is at least 0.5 in every pixel. This demonstrates a significant increase in the information content of every pixel due to the high mean gain multiplication. Moreover, the pixels in the center of the image have  $\alpha_{GeomR}$  values that are closer to 0.5, while the rest have  $\alpha_{GeomR}$  values that are closer to 1. This is in accordance with the curve with the same high mean gain in Fig. 3b which, though shown for  $\sigma_w = 8 e^-$ , indicates that the value of  $\alpha_{GeomR}$  decreases with increasing values of  $\nu_\theta$ . The minimum  $\alpha_{GeomR}$  value in this case is 0.503, which implies, by Theorem 7, that the Fisher information for this scenario is greater than half of that for the ideal scenario.

6.2.2 Limit of the localization accuracy comparison

To arrive at the limit of the localization accuracy for each data model, the estimated parameters are defined as the coordinates of the point source, i.e.,  $\theta = (x_0, y_0)$ . Given  $\theta$  and the noise coefficients computed above, we readily obtain the Fisher information matrix for each data model as described in Sect. 6.1. The limit of the localization accuracy for  $x_0$  ( $y_0$ ) is then defined as the square root of the Cramer-Rao lower bound on the variance of the estimates of  $x_0$  ( $y_0$ ), and is thus a lower bound on the standard deviation of any unbiased estimator of  $x_0$  ( $y_0$ ). Due to the centering of the circularly symmetric Airy image function on a square pixel array, the accuracy limits for  $x_0$  and  $y_0$  are the same. Hence, we give only a single limit for each data model.

The limits of the localization accuracy for the three data models are shown in Table 2. As expected, the ideal scenario of Poisson-distributed data has the best (i.e., the smallest) accuracy limit of 8.50 nm. In comparison, the CCD scenario has a significantly higher accuracy limit of 21.14 nm due to the readout noise of standard deviation  $\sigma_w = 8 e^-$  in every pixel. In contrast, by using high mean gain multiplication to drown out the readout noise, the EMCCD scenario has an accuracy limit of 11.58 nm that is only a little higher than that of the ideal scenario. Note that this is despite a larger readout noise standard deviation of  $\sigma_w = 24 e^-$  per pixel, which for the CCD scenario would more than double its accuracy limit to 57.27 nm.

Table 2 also shows, for each data model, the mean and standard deviation of the estimates of the  $x_0$  coordinate from maximum likelihood estimations carried out on 1000 simulated images of the point source. The image simulation and the estimation are performed in MATLAB (The MathWorks, Inc., Natick, MA). For an ideal image, the  $k$ th pixel value, with mean  $\nu_{\theta,k}$

**Table 2** Limits of the localization accuracy and results of maximum likelihood estimations for the ideal Poisson-distributed data model and the practical CCD and EMCCD image models

Data model	No. of $x_0$ estimates	True $x_0$ (nm)	Mean of $x_0$ estimates (nm)	Limit of the localization accuracy (nm)	Standard deviation of $x_0$ estimates (nm)
Ideal	1000	880	880.16	8.50	8.69
CCD	998 <sup>a</sup>	880	880.47	21.14	20.94
EMCCD	1000	880	879.74	11.58	11.75

Attributes of the imaged point source and all imaging-related parameters are as in Fig. 4. For each data model, the positional coordinates  $(x_0, y_0)$  of the point source are estimated from 1000 simulated images, and the results for  $x_0$  are shown

<sup>a</sup> Two outlier estimates which place the center of the point source’s image outside the pixel array have been removed

of Eq. 29, is generated using MATLAB's *poissrnd* function. For a CCD image, the  $k$ th pixel value is the sum of a similarly generated Poisson random number and a Gaussian random number with mean 0 and standard deviation 8 generated using MATLAB's *randn* function. For an EMCCD image, the  $k$ th pixel value is the sum of a number generated from the standard geometric multiplication probability mass function of Eq. 12 (with  $m^N = 1015.46$  and  $v_\theta = v_{\theta,k}$ ) using the inverse transform method, and a Gaussian random number with mean 0 and standard deviation 24.

By the independence of the pixels, the log-likelihood function that is maximized for a given image is the sum of the log-likelihood functions of its individual pixels. For an ideal and a CCD image, the log-likelihood function of the  $k$ th pixel is the logarithm of, respectively, the Poisson mass function with mean  $v_{\theta,k}$  of Eq. 29 and the density function given in the statement of Corollary 2 with  $v_\theta = v_{\theta,k}$ ,  $\eta_w = 0$ , and  $\sigma_w = 8$ . For an EMCCD image, the  $k$ th pixel's log-likelihood function is approximated by the density function of Eq. 41 with  $m^N = 1015.46$ ,  $v_\theta = v_{\theta,k}$ ,  $\eta_w = 0$ , and  $\sigma_w = 24$ . The high gain approximation is used (as opposed to the standard geometric multiplication density function of Eq. 37) because it computes significantly faster while, as the results will show, retaining reasonable accuracy. In all cases, the maximization is realized by minimizing the negative of the log-likelihood function using MATLAB's *fminunc* function. Also, initial estimates for  $x_0$  and  $y_0$  are randomly generated per image to be within 20% of the true value of 880 nm.

For each data scenario, Table 2 shows that the standard deviation of the maximum likelihood estimates of  $x_0$  comes reasonably close to the corresponding limit of the localization accuracy. In each case, the mean of the estimates also recovers reasonably closely the true value of  $x_0$ . (For the results of maximum likelihood estimations performed on additional simulated EMCCD data sets that differ in the expected number of detected photons  $N_{\text{photon}}$ , see Table 3 in the Appendix. As in the case of the results of Table 2, the mean values and the standard deviations of the  $x_0$  estimates of these data sets reasonably recover, respectively, the true value of  $x_0$  and their corresponding limits of the localization accuracy.)

## 7 Conclusion

We have derived various Fisher information expressions pertaining to data that can be described as the output of a branching process with potentially added measurement noise. All expressions assume a Poisson-distributed initial signal which is characteristic of many physical processes. General expressions which make no assumptions about the model of signal amplification allow the straightforward derivation of specific expressions once the probability distribution of the amplified signal is known. On the other hand, specific results which assume the geometric model of amplification can readily be applied to, for example, imaging with an EMCCD. Meanwhile, expressions based on two approximations of an amplified signal offer computational efficiency and/or ease of analysis. Throughout the presentation, illustration and comparison of different data models have been performed using a noise coefficient that has been defined based on the common form taken by the data models. As demonstrated with a point source localization example, the developed theory has applications in the Fisher information analysis of practically important problems.

**Acknowledgments** This work was supported in part by grants from the National Institutes of Health (R01 GM071048 and R01 GM085575).

**Appendix**

*Proof of Corollary 2:* Since  $p_{\theta,R}$  is parameterized by  $\theta$  through  $v_\theta$ , by Theorem 1  $\mathbf{I}_R(\theta)$  takes the form of Eq. 1. To prove the result, we need only evaluate the expectation term of Eq. 1 with  $p_\theta = p_{\theta,R}$ :

$$\begin{aligned}
 E \left[ \left( \frac{\partial}{\partial v_\theta} \ln(p_{\theta,R}(z)) \right)^2 \right] &= E \left[ \left( \frac{1}{p_{\theta,R}(z)} \frac{e^{-v_\theta}}{\sqrt{2\pi}\sigma_w} \sum_{j=0}^\infty \frac{j v_\theta^{j-1} - v_\theta^j}{j!} e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} \right)^2 \right] \\
 &= E \left[ \left( \frac{1}{p_{\theta,R}(z)} \left( \frac{e^{-v_\theta}}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \frac{v_\theta^{j-1}}{(j-1)!} e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} - p_{\theta,R}(z) \right) \right)^2 \right] \\
 &= E \left[ \left( \frac{1}{p_{\theta,R}(z)} \frac{e^{-v_\theta}}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \frac{v_\theta^{j-1}}{(j-1)!} e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} \right)^2 \right] \\
 &\quad - 2 \cdot E \left[ \frac{1}{p_{\theta,R}(z)} \frac{e^{-v_\theta}}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \frac{v_\theta^{j-1}}{(j-1)!} e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} \right] + 1 \\
 &= \int_{-\infty}^\infty \frac{1}{p_{\theta,R}(z)} \left( \frac{e^{-v_\theta}}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \frac{v_\theta^{j-1}}{(j-1)!} e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} \right)^2 dz \\
 &\quad - \frac{2 \cdot e^{-v_\theta}}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \left( \frac{v_\theta^{j-1}}{(j-1)!} \int_{-\infty}^\infty e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} dz \right) + 1 \\
 &= \int_{-\infty}^\infty \frac{1}{p_{\theta,R}(z)} \left( \frac{e^{-v_\theta}}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \frac{v_\theta^{j-1}}{(j-1)!} e^{-\frac{1}{2} \left( \frac{z-j-\eta_w}{\sigma_w} \right)^2} \right)^2 dz - 1. \quad \square
 \end{aligned}$$

**Corollary 4** Let  $Z_\theta = X_{N,\theta} + W$ , where  $X_{N,\theta}$  is as defined in Theorem 3, and  $W$  is a Gaussian random variable with mean  $\eta_w$  and variance  $\sigma_w^2$ . Let  $X_{N,\theta}$  and  $W$  be stochastically independent of each other, and let  $W$  be not dependent on  $\theta$ .

1. The probability distribution of  $Z_\theta$  is given by the probability density function

$$p_{\theta,MR}(z) = \frac{1}{\sqrt{2\pi}\sigma_w} \sum_{l=0}^\infty e^{-\left(\frac{z-l-\eta_w}{\sqrt{2}\sigma_w}\right)^2} \left[ \sum_{j=0}^\infty P(X_{N,\theta} = l | X_{0,\theta} = j) \cdot \frac{e^{-v_\theta} v_\theta^j}{j!} \right], \quad z \in \mathbb{R}. \tag{31}$$

2. The noise coefficient corresponding to the probability density  $p_{\theta,MR}$  is given by

$$\alpha_{MR} = \int_{-\infty}^\infty \frac{v_\theta \cdot e^{-2v_\theta}}{p_{\theta,MR}(z)} \left( \sum_{l=0}^\infty \frac{e^{-\left(\frac{z-l-\eta_w}{\sqrt{2}\sigma_w}\right)^2}}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^\infty \frac{P(X_{N,\theta} = l | X_{0,\theta} = j) \cdot v_\theta^{j-1}}{(j-1)!} \right)^2 dz - v_\theta. \tag{32}$$

The Fisher information matrix of  $Z_\theta$  is given by  $\mathbf{I}_{MR}(\theta) = \alpha_{MR} \cdot \mathbf{I}_P(\theta)$ , with  $\mathbf{I}_P(\theta)$  as given in Eq. 2.

*Proof of Theorem 4* 1. Let  $f$  be the probability generating function of the zero modified geometric distribution  $(p_k)_{k=0,1,\dots}$  of Eq. 9. By the definition of the generating function, for  $s \in \mathbb{C}$  and  $|s| \leq 1$ ,

$$f(s) = \sum_{k=0}^{\infty} p_k s^k = p_0 s^0 + \sum_{k=1}^{\infty} p_k s^k = a + \sum_{k=1}^{\infty} (1-a)(1-b)b^{k-1} s^k$$

$$= \frac{a - (a - (1-b))s}{1-bs}.$$

Given an  $N$ -stage branching process with initial particle count 1 (i.e.,  $U = 1$  in Definition 2) and an individual offspring count probability distribution  $(p_k)_{k=0,1,\dots}$ , the probability generating function of the probability distribution of the particle count  $X_{N,\theta}$  at the output of the branching process is well known to be the  $N$ th iterate of the probability generating function of  $(p_k)_{k=0,1,\dots}$  (e.g., Harris 1963; Athreya and Ney 2004; Grimmett and Stirzaker 2001). Denoted by  $f_N(s)$ , the  $N$ th iterate of  $f(s)$  can be shown by induction to be, for  $m = \frac{1-a}{1-b} \neq 1$ ,

$$f_N(s) = 1 - \frac{(1-a)(m-1)m^N(1-s)}{b(m^N-1)m(1-s) + (1-a)(m-1)}, \quad N = 0, 1, \dots, \quad s \in \mathbb{C}, \quad |s| \leq 1. \tag{33}$$

According to Definition 2, each initial particle of a branching process multiplies independently of the other initial particles, but according to a common stochastic model. Therefore, if the initial particle count is a positive integer  $j$ , then the probability mass function of the output particle count  $X_{N,\theta}$  will be the  $j$ -fold convolution of, in this case, the probability mass function corresponding to Eq. 33 with itself. By a well known result from the theory of generating functions, the generating function of  $X_{N,\theta}$  is then just  $[f_N(s)]^j$  (e.g., Grimmett and Stirzaker 2001).

Using this generating function, the conditional probability that the output particle count  $X_{N,\theta} = x$  given  $j \in \{1, 2, \dots\}$  initial particles can be obtained based on the definition of the generating function. For  $x = 0$ , it is given by  $[f_N(s)]^j$  evaluated at  $s = 0$ , and for  $x = 1, 2, \dots$ , it is given by the  $x$ th derivative of  $[f_N(s)]^j$  evaluated at  $s = 0$  and divided by  $x!$  (e.g., Grimmett and Stirzaker 2001). This can be shown to yield, for  $A, B, C$ , and  $D$  as given for Eq. 10,

$$P(X_{N,\theta} = x | X_{0,\theta} = j)$$

$$= \begin{cases} \left[1 - \frac{A}{B}\right]^j, & x = 0, \\ \sum_{l=0}^{x-1} \frac{\binom{x-1}{l} C^{x-1-l} D^{l+1}}{(l+1)! B^{x+l+1}} \frac{\partial^{l+1}}{\partial \left(1 - \frac{A}{B}\right)^{l+1}} \left(\left[1 - \frac{A}{B}\right]^j\right), & x = 1, 2, \dots \end{cases} \tag{34}$$

Note that Eq. 34 also applies for  $j = 0$ , since  $P(X_{N,\theta} = x | X_{0,\theta} = 0)$  is 1 for  $x = 0$ , and is 0 otherwise. Finally, for an initial particle count that is Poisson-distributed with mean  $\nu_\theta$ , using the conditional mass function of Eq. 34 in the general expression of Eq. 7 gives the desired probability mass function of  $X_{N,\theta}$ .

2. The noise coefficient  $\alpha_{Geom}$  follows by substituting Eqs. 34 and 10 into Eq. 8 and simplifying. The Fisher information matrix  $\mathbf{I}_{Geom}(\theta)$  then follows directly from Definition 1. □

**Corollary 5** Let  $Z_\theta = X_{N,\theta} + W$ , where  $X_{N,\theta}$  is as defined in Theorem 4, and  $W$  is a Gaussian random variable with mean  $\eta_w$  and variance  $\sigma_w^2$ . Let  $X_{N,\theta}$  and  $W$  be stochastically independent of each other, and let  $W$  be not dependent on  $\theta$ .

1. The probability density function of  $Z_\theta$  is, for  $z \in \mathbb{R}$ , given by

$$p_{\theta, \text{GeomR}}(z) = \frac{e^{-v_\theta \frac{A}{B}}}{\sqrt{2\pi} \sigma_w} \left[ e^{-\left(\frac{z-\eta_w}{\sqrt{2}\sigma_w}\right)^2} + \sum_{l=1}^{\infty} e^{-\left(\frac{z-l-\eta_w}{\sqrt{2}\sigma_w}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j} C^{l-1-j} (Dv_\theta)^{j+1}}{(j+1)! B^{j+l+1}} \right]. \tag{35}$$

2. The noise coefficient corresponding to the probability density  $p_{\theta, \text{GeomR}}$  is given by

$$\alpha_{\text{GeomR}} = \int_{-\infty}^{\infty} \frac{v_\theta D^2 e^{-2v_\theta \frac{A}{B}}}{p_{\theta, \text{GeomR}}(z)} \left( \sum_{l=1}^{\infty} e^{-\left(\frac{z-l-\eta_w}{\sqrt{2}\sigma_w}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j} C^{l-1-j} (Dv_\theta)^{-j}}{j! B^{j+l+1}} \right)^2 dz - v_\theta \frac{A^2}{B^2}. \tag{36}$$

The Fisher information matrix of  $Z_\theta$  is  $\mathbf{I}_{\text{GeomR}}(\theta) = \alpha_{\text{GeomR}} \cdot \mathbf{I}_P(\theta)$ , with  $\mathbf{I}_P(\theta)$  as given in Eq. 2.

For standard geometric multiplication, setting  $a = 0$  in Eqs. 35 and 36 yields, for  $m = \frac{1}{1-b}$ ,

$$p_{\theta, \text{GeomR}}(z) = \frac{e^{-v_\theta}}{\sqrt{2\pi} \sigma_w} \left[ e^{-\left(\frac{z-\eta_w}{\sqrt{2}\sigma_w}\right)^2} + \sum_{l=1}^{\infty} e^{-\left(\frac{z-l-\eta_w}{\sqrt{2}\sigma_w}\right)^2} \times \sum_{j=0}^{l-1} \frac{\binom{l-1}{j} \left(1 - \frac{1}{m^N}\right)^{l-j-1}}{(j+1)! \left(\frac{m^N}{v_\theta}\right)^{j+1}} \right], \quad z \in \mathbb{R}, \tag{37}$$

$$\alpha_{\text{GeomR}} = \int_{-\infty}^{\infty} \frac{v_\theta \cdot e^{-2v_\theta}}{p_{\theta, \text{GeomR}}(z)} \left( \sum_{l=1}^{\infty} e^{-\left(\frac{z-l-\eta_w}{\sqrt{2}\sigma_w}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j} \left(1 - \frac{1}{m^N}\right)^{l-j-1}}{j! \left(\frac{m^N}{v_\theta}\right)^j} \right)^2 dz - v_\theta. \tag{38}$$

**Corollary 6** Let  $Z_\theta = X_\theta + W$ , where  $X_\theta$  is as defined in Theorem 5, and  $W$  is a Gaussian random variable with mean  $\eta_w$  and variance  $\sigma_w^2$ . Let  $X_\theta$  and  $W$  be stochastically independent of each other, and let  $W$  be not dependent on  $\theta$ .

1. The probability density function of  $Z_\theta$  is, for  $0 \leq a < b < 1$  and  $m = \frac{1-a}{1-b}$ ,

$$p_{\theta, \text{HR}}(z) = \frac{1}{\sqrt{2\pi} \sigma_w} \left[ \frac{e^{-\left(\frac{z-\eta_w}{\sqrt{2}\sigma_w}\right)^2}}{e^{v_\theta(1-\frac{a}{b})}} + \frac{\left(1 - \frac{a}{b}\right)^2}{m^N} \times \int_0^\infty \frac{e^{-\left(\frac{z-u-\eta_w}{\sqrt{2}\sigma_w}\right)^2}}{e^{\frac{u}{m^N} + v_\theta}} \sum_{j=1}^{\infty} \frac{v_\theta^j}{(j-1)!} \sum_{k=0}^{j-1} \frac{\binom{j-1}{k} \left(\frac{a}{b}\right)^{j-k-1}}{(k+1)! \left(\left(1 - \frac{a}{b}\right)^2 \frac{u}{m^N}\right)^{-k}} du \right]. \tag{39}$$

**Table 3** Limits of the localization accuracy and results of maximum likelihood estimations for four simulated data sets of the EMCCD image model

Expected photon count $N_{photon}$	True $x_0$ (nm)	Mean of $x_0$ estimates (nm)	Limit of the localization accuracy (nm)	Standard deviation of $x_0$ estimates (nm)	Ideal limit of the localization accuracy (nm)	CCD limit of the localization accuracy (nm)
150	880	879.56	13.28	13.19	9.82	27.44
100	880	879.80	16.07	16.18	12.02	39.99
50	880	880.45	21.98	22.23	17.00	77.53
25	880	879.35	29.61	31.32	24.04	152.53

Except for the expected number of detected photons  $N_{photon}$ , for each data set the attributes of the imaged point source and all imaging-related parameters are as in Fig. 4b. For each data set, the positional coordinates  $(x_0, y_0)$  of the point source are estimated from 1000 simulated images, and the results obtained from the 1000  $x_0$  estimates are shown. In addition, limits of the localization accuracy for the corresponding ideal Poisson-distributed data model and CCD image model (with readout noise parameters as given in Fig. 4a) are provided for comparison

2. The noise coefficient corresponding to the probability density  $p_{\theta,HR}$  is given by

$$\alpha_{HR} = \int_{-\infty}^{\infty} \frac{v_{\theta}}{2\pi\sigma_w^2 \cdot p_{\theta,HR}(z)} \left( \frac{a \cdot e^{-\left(\frac{z-\eta w}{\sqrt{2}\sigma_w}\right)^2}}{b \cdot e^{v_{\theta}\left(1-\frac{a}{b}\right)}} + \frac{\left(1-\frac{a}{b}\right)^2}{m^N} \right) \times \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta w}{\sqrt{2}\sigma_w}\right)^2}}{e^{\left(1-\frac{a}{b}\right)\frac{u}{m^N} + v_{\theta}}} \sum_{j=1}^{\infty} \frac{j v_{\theta}^{j-1}}{(j-1)!} \sum_{k=0}^{j-1} \frac{(j-1) \left(\frac{a}{b}\right)^{j-k-1}}{(k+1)! \left(\left(1-\frac{a}{b}\right)^2 \frac{u}{m^N}\right)^{-k}} du \right)^2 dz - v_{\theta}. \tag{40}$$

The Fisher information matrix of  $Z_{\theta}$  is given by  $\mathbf{I}_{HR}(\theta) = \alpha_{HR} \cdot \mathbf{I}_P(\theta)$ , where  $\mathbf{I}_P(\theta)$  is as in Eq. 2.

To approximate standard geometric multiplication, setting  $a = 0$  in Eqs. 39 and 40 yields, for  $m = \frac{1}{1-b}$ ,  $0 < b < 1$ ,

$$p_{\theta,HR}(z) = \frac{e^{-v_{\theta}}}{\sqrt{2\pi}\sigma_w} \left[ e^{-\left(\frac{z-\eta w}{\sqrt{2}\sigma_w}\right)^2} + \int_0^{\infty} e^{-\left(\frac{z-u-\eta w}{\sqrt{2}\sigma_w}\right)^2} - \frac{u}{m^N} \frac{\sqrt{v_{\theta}u}}{m^N} I_1\left(2\sqrt{\frac{v_{\theta}u}{m^N}}\right) du \right], \quad z \in \mathbb{R}, \tag{41}$$

$$\alpha_{HR} = v_{\theta} \cdot \left[ \int_{-\infty}^{\infty} \frac{1}{p_{\theta,HR}(z)} \left( \frac{e^{-v_{\theta}}}{\sqrt{2\pi}\sigma_w} \int_0^{\infty} \frac{e^{-\frac{1}{2}\left(\frac{z-u-\eta w}{\sigma_w}\right)^2} - \frac{u}{m^N}}}{m^N} I_0\left(2\sqrt{\frac{v_{\theta}u}{m^N}}\right) du \right)^2 dz - 1 \right]. \tag{42}$$

**References**

Abramowitz, M. & Stegun, I. A. (Eds.). (1965). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover Publications.  
 Andersson, S. B. (2008). Localization of a fluorescent source without numerical fitting. *Optics Express*, 16, 18714–18724.  
 Athreya, K. B., & Ney, P. E. (2004). *Branching processes*. New York: Dover Publications.

- Basden, A. G., Haniff, C. A., & Mackay, C. D. (2003). Photon counting strategies with low-light-level CCDs. *Monthly Notices of the Royal Astronomical Society*, *345*, 985–991.
- Born, M., & Wolf, E. (1999). *Principles of optics*. Cambridge: Cambridge University Press.
- DeSantis, M. C., DeCenzo, S. H., Li, J., & Wang, Y. M. (2010). Precision analysis for standard deviation measurements of immobile single fluorescent molecule images. *Optics Express*, *18*, 6563–6576.
- Grimmett, G. R., & Stirzaker, D. R. (2001). *Probability and random processes* (3rd ed.). New York: Oxford University Press.
- Haccou, P., Jagers, P., & Vatutin, V. A. (2005). *Branching processes: Variation, growth, and extinction of populations*. Cambridge: Cambridge University Press.
- Harris, T. E. (1963). *The theory of branching processes*. Berlin: Springer.
- Hollenhorst, J. N. (1990). A theory of multiplication noise. *IEEE Transactions on Electron Devices*, *37*, 781–788.
- Hynecek, J., & Nishiwaki, T. (2003). Excess noise and other important characteristics of low light level imaging using charge multiplying CCDs. *IEEE Transactions on Electron Devices*, *50*, 239–245.
- Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. New Jersey: Prentice Hall PTR.
- Kimmel, M., & Axelrod, D. E. (2002). *Branching processes in biology*. New York: Springer.
- Matsuo, K., Teich, M. C., & Saleh, B. E. A. (1985). Noise properties and time response of the staircase avalanche photodiode. *IEEE Transactions on Electron Devices*, *ED-32*, 2615–2623.
- Moerner, W. E. (2007). New directions in single-molecule imaging and analysis. *Proceedings of the National Academy of Sciences USA*, *104*, 12596–12602.
- Mortensen, K. I., Churchman, L. S., Spudich, J. A., & Flyvbjerg, H. (2010). Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nature Methods*, *7*, 377–381.
- Ober, R. J., Ram, S., & Ward, E. S. (2004). Localization accuracy in single-molecule microscopy. *Biophysical Journal*, *86*, 1185–1200.
- Pavani, S. R. P., & Piestun, R. (2008). Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system. *Optics Express*, *16*, 22048–22057.
- Quan, T., Zeng, S., & Huang, Z. (2010). Localization capability and limitation of electron-multiplying charge-coupled, scientific complementary metal-oxide semiconductor, and charge-coupled devices for superresolution imaging. *Journal of Biomedical Optics*, *15*, 066005.
- Ram, S., Ward, E. S., & Ober, R. J. (2006). A stochastic analysis of performance limits for optical microscopes. *Multidimensional Systems and Signal Processing*, *17*, 27–57.
- Ram, S., Prabhat, P., Chao, J., Ward, E. S., & Ober, R. J. (2008). High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells. *Biophysical Journal*, *95*, 6025–6043.
- Rao, C. R. (1965). *Linear statistical inference and its applications*. New York: Wiley.
- Robbins, M. S., & Hadwen, B. J. (2003). The noise performance of electron multiplying charge-coupled devices. *IEEE Transactions on Electron Devices*, *50*, 1227–1232.
- Snyder, D. L., Helstrom, C. W., Lanterman, A. D., Faisal, M., & White, R. L. (1995). Compensation for readout noise in CCD images. *Journal of the Optical Society of America A*, *12*, 272–283.
- Thompson, M. A., Lew, M. D., Badieirostami, M., & Moerner, W. E. (2010). Localizing and tracking single nanoscale emitters in three dimensions with high spatiotemporal resolution using a double-helix point spread function. *Nano Letters*, *10*, 211–218.
- Thompson, R. E., Larson, D. R., & Webb, W. W. (2002). Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal*, *82*, 2775–2783.
- Tubbs, R. N. (2003). *Lucky exposures: Diffraction limited astronomical imaging through the atmosphere*. PhD thesis, University of Cambridge.
- Walter, N. G., Huang, C., Manzo, A. J., & Sobhy, M. A. (2008). Do-it-yourself guide: How to use the modern single-molecule toolkit. *Nature Methods*, *5*, 475–489.
- Wu, P., Nelson, N., & Tseng, Y. (2010). A general method for improving spatial resolution by optimization of electron multiplication in CCD imaging. *Optics Express*, *18*, 5199–5212.

### Author Biographies



**Jerry Chao** obtained his B.S. and M.S. degrees in computer science from the University of Texas at Dallas, Richardson, in 2000 and 2002, respectively. He received the Ph.D. degree in electrical engineering from the same institution in 2010. From 2003 to 2005, he developed software for microscopy image acquisition and analysis at the University of Texas Southwestern Medical Center, Dallas. He is currently carrying out postdoctoral research in the Department of Electrical Engineering, University of Texas at Dallas, and the Department of Immunology, University of Texas Southwestern Medical Center. His research interests include biomedical image and signal processing, with particular application to cellular microscopy, and the development of software for bioengineering applications.



**E. Sally Ward** received the Ph.D. degree from the Department of Biochemistry, Cambridge University, Cambridge, U.K., in 1985. From 1985 to 1987, she was a Research Fellow at Gonville and Caius College while working at the Department of Biochemistry, Cambridge University. From 1988 to 1990, she held the Stanley Elmore Senior Research Fellowship at Sidney Sussex College and carried out research at the MRC Laboratory of Molecular Biology, Cambridge. In 1990, she joined the University of Texas Southwestern Medical Center, Dallas, as an Assistant Professor. Since 2002, she has been a Professor in the Department of Immunology at the same institution, and currently holds the Paul and Betty Meek-FINA Professorship in Molecular Immunology. Her research interests include antibody engineering, molecular mechanisms that lead to autoimmune disease, questions related to the in vivo dynamics of antibodies, and the use of microscopy techniques for the study of antibody trafficking in cells.



**Raimund J. Ober** received the Ph.D. degree in engineering from Cambridge University, Cambridge, U.K., in 1987. From 1987 to 1990, he was a Research Fellow at Girton College and the Engineering Department, Cambridge University. In 1990, he joined the University of Texas at Dallas, Richardson, where he is currently a Professor with the Department of Electrical Engineering. He is also Adjunct Professor at the University of Texas Southwestern Medical Center, Dallas. He is an Associate Editor of *Mathematics of Control, Signals, and Systems*, and a past Associate Editor of *Multidimensional Systems and Signal Processing*, *IEEE Transactions on Circuits and Systems*, and *Systems and Control Letters*. His research interests include the development of microscopy techniques for cellular investigations, in particular at the single molecule level, the study of cellular trafficking pathways using microscopy investigations, and signal/image processing of bioengineering data.