Claremont Colleges
Scholarship @ Claremont

All HMC Faculty Publications and Research

HMC Faculty Scholarship

4-1-1985

Problems of Channel Correlation and Statistical Bias in Photon-Correlation Spectroscopy

Richard C. Haskell
Harvey Mudd College

Gary L. Pisciotta

Problems of channel correlation and statistical bias in photon-correlation spectroscopy

Richard C. Haskell and Gary L. Pisciotta*

Department of Physics, Harvey Mudd College, Claremont, California 91711

Correlation between channels of the normalized photocount-rate correlation function $g^{(2)}(\tau)$ becomes significant at high count rates and leads to a number of data-analysis problems. We derive an expression for channel correlation that is valid for a detector area of arbitrary extent and compare the theoretical predictions with measured values. A data-analysis procedure is demonstrated that employs the theoretical expression for channel correlation and provides a rigorous test of an assumed fitting function. The procedure facilitates the use of the cumulant method in determining the polydispersity of scatterers. An expression for the statistical bias of $g^{(2)}(\tau)$ is also derived and compared with measured values.

1. INTRODUCTION

The light-scattering technique of photon-correlation spectroscopy (PCS) is now 20 years old[1,2] and is employed in a wide range of research programs in which frequency shifts of less than 1 MHz are of interest. In PCS a measurement is performed of the normalized photocount-rate correlation function $g^{(2)}(\tau)$ of light scattered from the sample. From the measured $g^{(2)}(\tau)$ one can deduce the dynamics of the motion of the scatterers or the time scale of polarizability fluctuations in the scattering medium. In an increasing number of PCS applications the average number of photocounts recorded in a coherence time of the scattered optical field is quite large ($\bar{n}_c \gg 1$). Examples of such applications include resting and contracting skeletal muscle,[3] macromolecules diffusing in solution (e.g., the muscle protein F-actin[4]), studies of cell motility,[5,6] and resting and contracting cardiac muscle.[7] Indeed, the condition $\bar{n}_c \gg 1$ is realized in the most common PCS calibration procedure in which laser light is scattered from a sample of uniform-diameter (e.g., 0.1-μm) spheres executing Brownian motion in water. In most of these situations the coherence time of the scattered optical field is so long that any attempt to reduce the photocount rate in order to achieve $\bar{n}_c \lesssim 1$ results in a photocount rate that is unsatisfactorily close to the dark-count rate of the photomultiplier. Hence the condition $\bar{n}_c \gg 1$ often cannot be avoided.

Unfortunately, when $\bar{n}_c \gg 1$ the channels of $g^{(2)}(\tau)$ [the measured values of $g^{(2)}(\tau)$ at different values of τ] are correlated, and two problems arise during data analysis. First, when the measured $g^{(2)}(\tau)$ is fitted to some expected functional form (e.g., a single exponential in the case of a Lorentzian light spectrum), a ridiculously low value for chi squared per degree of freedom (χ^2/DF) is found. Second, the F-test criterion frequently employed with the cumulant technique[8] for determining the polydispersity of the scatterers is rendered invalid by the erroneous values for χ^2/DF.

Saleh and Cardoso[9] have investigated the problem of channel correlation in the case of a stationary, cross-spectrally pure, Gaussian scattered field with arbitrary spectral line shape. Their treatment assumes that the detector area is small compared with the coherence area of the scattered field, that the sample time [width of a channel of $g^{(2)}(\tau)$] is short compared with the coherence time τ_c, and that the duration of a measurement of $g^{(2)}(\tau)$ (correlator run duration) is long compared with τ_c. They obtained an explicit expression for channel correlation in the frequently occurring case of a Lorentzian spectrum and estimated the statistical accuracy of a measurement of the spectral parameter $\tau_c{}^{-1}$ for various photocount rates and measurement times. In the present paper we shall be interested not only in the accuracy with which regression techniques can extract a spectral parameter but also in the values obtained for χ^2/DF that indicate whether the assumed spectrum is appropriate and in the results of F tests that help to determine the polydispersity of the scatterers. We shall extend the treatment of Saleh and Cardoso[9] to the case of an arbitrary detector area and shall compare our calculated expression for channel correlation with values measured by scattering laser light from latex spheres in water. We shall then suggest a simple on-line technique for taking account of channel correlation and for removing the two data-analysis problems mentioned above.

In order to calculate the channel correlation we must first discuss the statistical bias of $g^{(2)}(\tau)$, which becomes appreciable when the duration of a measurement of $g^{(2)}(\tau)$ is less than approximately 200 coherence times. Since the time available for measurement can be quite limited in PCS measurements performed on biological systems,[3] the bias itself can be of considerable interest. In Section 2 we calculate the bias of $g^{(2)}(\tau)$ and compare the theoretical expression with values measured by scattering laser light from latex spheres in ethylene glycol. Section 3 contains a calculation of the covariance of two channels of $g^{(2)}(\tau)$ and a comparison of the theoretical predictions with measured values. In Section 4 the theoretical expression for the covariance is incorporated into a procedure for fitting measured values of $g^{(2)}(\tau)$ to some expected functional form. The procedure is demonstrated by analyzing data obtained with a range of photocount rates, illustrating the transition from correlated channels ($\bar{n}_c \gg 1$) to uncorrelated channels ($\bar{n}_c \lesssim 1$). A summary and discussion of results are presented in Section 5.

2. STATISTICAL BIAS OF $g^{(2)}(\tau)$

In order to perform a measurement of $g^{(2)}(\tau)$, a digital correlator divides a real-time interval called the correlator run duration into a large number N of contiguous time intervals of duration T called sample times. The lth channel of the normalized photocount-rate correlation function is then calculated according to

$$g^{(2)}(\tau = lT) = \left[(1/N)\sum_{m=1}^{N}n(m)n(m+l)\right]\Big/$$
$$\left[(1/N)\sum_{m=1}^{N}n(m)\right]^{2}, \quad l = 1, 2, 3, \ldots, L, \tag{1}$$

where $n(m)$ is the number of photocounts recorded in the mth sample time and L is the number of channels calculated by the correlator. Following the approach of Jakeman et al.[10] and of Saleh and Cardoso,[9] we define the random variables

$$\hat{n} = (1/N)\sum_{m=1}^{N}n(m), \tag{2}$$

$$\hat{G}_{l} = (1/N)\sum_{m=1}^{N}n(m)n(m+l), \tag{3}$$

$$\hat{g}_{l} = \hat{G}_{l}/(\hat{n})^{2}, \tag{4}$$

where $\hat{n}$ is the estimator of the average number of photocounts recorded in a sample time and $\hat{G}_{l}$ and $\hat{g}_{l}$ are the estimators of the unnormalized and normalized photocount-rate correlation functions, respectively. A single correlator run yields one value for each of these three random variables. The probability distributions of $\hat{n}$, $\hat{G}_{l}$, and $\hat{g}_{l}$ can be calculated from the results of a large number of statistically independent correlator runs. Since the average of the ratio of two random variables does not in general equal the ratio of the averages, $\hat{g}_{l}$ is a biased estimator, and we define

$$\langle \hat{g}_{l}\rangle = \langle \hat{G}_{l}/(\hat{n})^{2}\rangle = \langle \hat{G}_{l}\rangle/\langle \hat{n}\rangle^{2} + \text{BIAS}_{l}, \tag{5}$$

where the angle brackets denote an average over many measurements (correlator runs). Introducing the notation $\langle \hat{n}\rangle = \bar{n}$, $\langle \hat{G}_{l}\rangle = \bar{G}_{l}$, and defining $g_{l} = \bar{G}_{l}/\bar{n}^{2}$, we can write

$$\langle \hat{g}_{l}\rangle = \bar{G}_{l}/\bar{n}^{2} + \text{BIAS}_{l} = g_{l} + \text{BIAS}_{l}. \tag{6}$$

Although the normalized estimator $\hat{g}_{l}$ is biased, measurements of $\hat{g}_{l}$ have a distinct advantage: $\hat{g}_{l}$ depends only on the spectral parameters of the scattered light, the detector area, and the sample time T and not on the photocount rate or the correlator run duration. For example, slow drifts in the incident laser intensity do not affect measurements of $\hat{g}_{l}$ if the correlator run duration (NT) is chosen to be short compared with the time scale of these drifts. However, if the correlator run duration is too short, the bias of $\hat{g}_{l}$ can become appreciable; this, if unrecognized, would lead to an erroneous interpretation of the spectrum. This last consideration is critical in situations in which the stochastic process underlying the dynamics of the scattering sample is stationary for limited periods of time. This is often true of biological samples.[3] It should also be noted that the particular method of normalization used in Eq. (1) has been shown to provide the best statistical accuracy in measurements of $g^{(2)}(\tau)$ and of the width of a Lorentzian spectrum.[11,12]

Since we shall need an expression for the bias of $\hat{g}_{l}$ in order

to calculate the channel correlation, we now calculate BIAS$_{l}$ as a function of the photocount rate and the correlator run duration. We shall employ the assumptions mentioned in Section 1: (1) The scattered optical field is stationary, cross-spectrally pure, and Gaussian and has a Lorentzian spectrum; (2) the number of coherence times N_{c} in a correlator run duration is large ($N_{c} \gg 1$); and (3) the coherence time is much longer than the sample time ($\tau_{c}/T \gg 1$). Under assumption (2) ($N_{c} \gg 1$), the number of statistically independent terms in the sums of Eqs. (2) and (3) is large. According to the central-limit theorem, one can assume the probability distribution of $\hat{G}_{l}$ and $\hat{n}$ to be a jointly Gaussian distribution. This assumption is weakest when N_{c} is smallest, which is when BIAS$_{l}$ is most appreciable. The shortest correlator run duration that we shall consider corresponds to $N_{c} = 20$, and we found that even in this worst case the measured distributions of $\hat{G}_{l}$ and $\hat{n}$ were approximately normal. With the assumption of a jointly Gaussian distribution for $\hat{G}_{l}$ and $\hat{n}$ we have

$$\langle \hat{g}_{l}\rangle = \langle \hat{G}_{l}/(\hat{n})^{2}\rangle = \iint (G_{l}/n^{2})P(G_{l}, n)\mathrm{d}G_{l}\mathrm{d}n, \tag{7}$$

where $P(G_{l}, n) = (2\pi)^{-1}|M|^{-1/2}\exp(-\tfrac{1}{2}v^{\dagger}M^{-1}v)$, v is the column vector

$$v = \begin{bmatrix}(G_{l} - \bar{G}_{l})\\(n - \bar{n})\end{bmatrix},$$

M is the covariance matrix of $\hat{G}_{l}$ and $\hat{n}$, i.e.,

$$M_{11} = \text{Var}(\hat{G}_{l}) = \langle(\hat{G}_{l} - \bar{G}_{l})^{2}\rangle,$$
$$M_{12} = M_{21} = \text{Cov}(\hat{G}_{l}, \hat{n}) = \langle(\hat{G}_{l} - \bar{G}_{l})(\hat{n} - \bar{n})\rangle,$$
$$M_{22} = \text{Var}(\hat{n}) = \langle(\hat{n} - \bar{n})^{2}\rangle,$$

and $|M|$ is the determinant of M. When the integration over G_{l} is performed, Eq. (7) becomes

$$\langle \hat{g}_{l}\rangle = [2\pi M_{22}]^{-1/2}\int \mathrm{d}n[(M_{12}/M_{22})(n - \bar{n})/\bar{n}^{2} + g_{l}]$$
$$\times [1 + (n - \bar{n})/\bar{n}]^{-2}\exp[-(n - \bar{n})^{2}/2M_{22}]. \tag{8}$$

The second factor in the integrand can be expanded in a binomial series:

$$[1 + (n - \bar{n})/\bar{n}]^{-2} = \sum_{m=0}^{\infty}(-1)^{m}(m + 1)[(n - \bar{n})/\bar{n}]^{m}, \tag{9}$$

with the result that the integral in Eq. (8) can be expressed as the sum of moments of $\hat{n}$. Equation (8) then become

$$\langle \hat{g}_{l}\rangle - g_{l} = \text{BIAS}_{l}$$
$$= \sum_{m=0}^{\infty}[(2m + 3)!!g_{l}\,\text{Var}(\hat{n})/\bar{n}^{2} - (2m + 1)!!$$
$$\times (m + 1)2\,\text{Cov}(\hat{G}_{l}, \hat{n})/\bar{n}^{3}][\text{Var}(\hat{n})/\bar{n}^{2}]^{m}, \tag{10}$$

where we use the notation $(2m + 1)!! = 1 \times 3 \times 5 \ldots (2m - 1)(2m + 1)$. For most cases of practical interest $\text{Var}(\hat{n})/\bar{n}^{2} < 10^{-3}$, and hence the first term ($m = 0$) in the asymptotic series of Eq. (10) provides a satisfactory expression for BIAS$_{l}$,

$$\text{BIAS}_{l}(m = 0) = 3g_{l}\,\text{Var}(\hat{n})/\bar{n}^{2} - 2\,\text{Cov}(\hat{G}_{l}, \bar{n})/\bar{n}^{3}. \tag{11}$$

This is the expression used by Saleh and Cardoso.[9] However, in order to illustrate the effect of the bias of $g^{(2)}(\tau)$, we shall consider correlator run durations as short as $20\,\tau_{c}$ for which $\text{Var}(\hat{n})/\bar{n}^{2} = 0.04$, and we find it necessary to keep terms
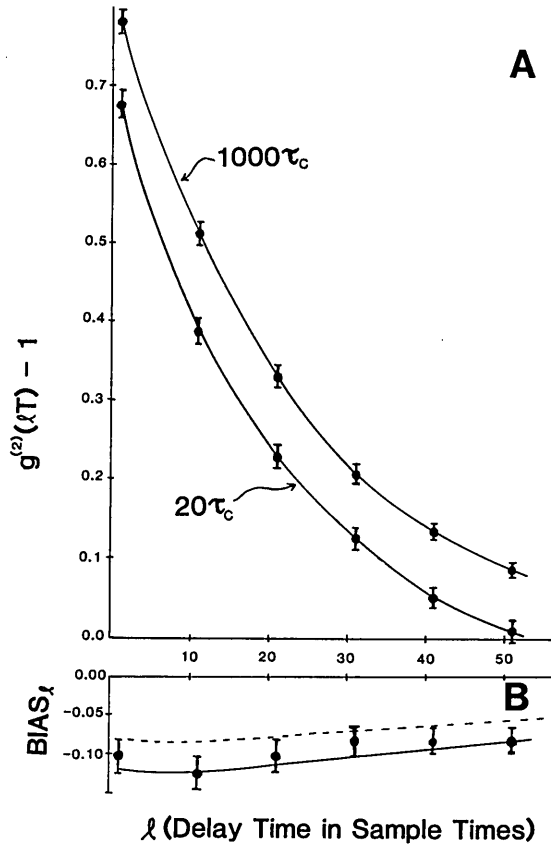
Fig. 1.   A, Measured values of $g^{(2)}(\tau) - 1$ are plotted for two values of the correlator-run duration: $NT = 20\,\tau_c$ and $NT = 1000\,\tau_c$. A He–Cd laser beam (10 mW, 441 nm) was scattered at 90° from 0.1-μm-diameter spheres in ethylene glycol at 23.8°C. The average photocount rate was 10,000 counts/sec. The sample time T was 0.1 msec, and the coherence time τ_c was 4.65 msec. The solid lines are drawn by hand through 50 data points. The error bars are ±1 standard deviation.   B, The data points are the measured bias of $g^{(2)}(\tau)$ for $N_c = 20$ calculated by taking the difference of the two curves in A. The solid line is the theoretical calculation for BIAS$_l$ with $N_c = 20$ keeping terms through $m = 4$ in Eq. (10). The dashed line represents the term for $m = 0$ only; see Eq. (11).

through $m = 4$ in Eq. (10) in order to obtain an accuracy of 1%.

In the case of a Lorentzian spectrum and for $N_c \gg 1 \gg \gamma = T/\tau_c$, we find the following expressions:

$$\mathrm{Var}(\hat{n})/\overline{n}^2 = N_c^{-1}(1/\overline{n}_c + a_2), \qquad (12)$$

$$\mathrm{Cov}(\hat{G}_l, \hat{n})/\overline{n}^3 = N_c^{-1}\{(2/\overline{n}_c)[1 + a_2\exp(-2\gamma l)]$$
$$+ 2a_2 + 2a_3(l\gamma + 1)\exp(-2\gamma l)\}. \qquad (13)$$

The a's account for integration of scattered light intensity over the area of the detector. For a point detector, the a's = 1; for a detector of finite size, $0 < a$'s < 1. A simple procedure for determining the spatial-integration factors is presented in Appendix A. For the experimental results presented in this paper, $a_2 = 0.817$, $a_3 = 0.740$, and $a_4 = 0.645$.

Using Eqs. (10), (12), and (13), we have calculated BIAS$_l$ for a range of correlator run durations and have plotted the results for $NT = 20\tau_c$ in Fig. 1B. Note that the solid line corresponds to the sum of terms through $m = 4$ in Eq. (10), whereas the dashed line corresponds to the $m = 0$ term only [Eq. (11)]. In order to check these theoretical predictions, we scattered the 10-mW beam of a He–Cd laser ($\lambda = 441$ nm) from 0.1-μm-diameter polystyrene latex spheres executing

Brownian motion in ethylene glycol. We measured $g^{(2)}(\tau)$ at a photocount rate corresponding to $\overline{n}_c = 50$ and for correlator run durations ranging from 20 τ_c to 20,000 τ_c. We found that the measured values of $g^{(2)}(\tau)$ for $NT \gtrsim 200\,\tau_c$ differed from those at $NT = 20{,}000\,\tau_c$ by less than 2%. In other words, the bias of $g^{(2)}(\tau)$ was insignificant for $NT \gtrsim 200\,\tau_c$. The measured values of $g^{(2)}(\tau)$ for $NT = 20\,\tau_c$ and $NT = 1000\,\tau_c$ are plotted in Fig. 1A, and the difference of the two curves, which should be the bias of $\hat{g}_l$ for $NT = 20\,\tau_c$, is plotted in Fig. 1B. The theoretical predictions and measured values are in reasonable agreement.

3.   COVARIANCE OF $\hat{g}_k$ AND $\hat{g}_l$

We now calculate the covariance of two channels of $g^{(2)}(\tau)$ using a procedure similar to that used in Section 2 to calculate the bias of $\hat{g}_l$. Guided again by the central-limit theorem, we assume that $\hat{G}_k$, $\hat{G}_l$, and $\hat{n}$ have a jointly Gaussian distribution. Therefore we can write

$$\langle \hat{g}_k\hat{g}_l\rangle = \langle \hat{G}_k\hat{G}_l/(\hat{n})^4\rangle = \int\!\!\int\!\!\int (G_kG_l/n^4)$$
$$\times P(G_k, G_l, n)\,dG_k\,dG_l\,dn, \qquad (14)$$

where $P(G_k, G_l, n) = (2\pi)^{-3/2}|M|^{-1/2}\exp(-\tfrac{1}{2}v^\dagger M^{-1}v)$, v is the column vector

$$v = \begin{bmatrix} (G_k - \overline{G}_k) \\ (G_l - \overline{G}_l) \\ (n - \overline{n}) \end{bmatrix},$$

and M is the covariance matrix, i.e.,

$$M_{11} = \mathrm{Var}(\hat{G}_k) = \langle (\hat{G}_k - \overline{G}_k)^2\rangle,$$

$$M_{12} = M_{21} = \mathrm{Cov}(\hat{G}_k, \hat{G}_l) = \langle (\hat{G}_k - \overline{G}_k)(\hat{G}_l - \overline{G}_l)\rangle,$$

$$M_{13} = M_{31} = \mathrm{Cov}(\hat{G}_k, \hat{n}) = \langle (\hat{G}_k - \overline{G}_k)(\hat{n} - \overline{n})\rangle,$$

etc., and $|M|$ is the determinant of M. The integration over G_k and G_l in Eq. (14) is lengthy but straightforward. The integration over n can be performed in a manner similar to that of Section 2 by expanding

$$[1 + (n - \overline{n})/\overline{n}]^{-4} = \sum_{m=0}^{\infty}(-1)^m(m + 1)(m + 2)$$
$$\times [(m + 3)/6][(n - \overline{n})/\overline{n}]^m. \qquad (15)$$

The final result of the integration in Eq. (14) is

$$\langle \hat{g}_k\hat{g}_l\rangle = [(\overline{n})^{-4}\,\mathrm{Cov}(\hat{G}_k, \hat{G}_l) + g_kg_l]$$

$$\times \sum_{m=0}^{\infty}[(2m + 3)!!(2m + 2)/6][\mathrm{Var}(\hat{n})/\overline{n}^2]^m$$

$$+ (\overline{n})^{-6}\,\mathrm{Cov}(\hat{G}_k, \hat{n})\mathrm{Cov}(\hat{G}_l, \hat{n})]$$

$$\times \sum_{m=0}^{\infty}[(2m + 5)!!(2m + 2)(2m + 4)/6][\mathrm{Var}(\hat{n})/\overline{n}^2]^m$$

$$- (\overline{n})^{-3}[g_k\,\mathrm{Cov}(\hat{G}_l, \hat{n}) + \mathrm{Cov}(\hat{G}_k, \hat{n})g_l]$$

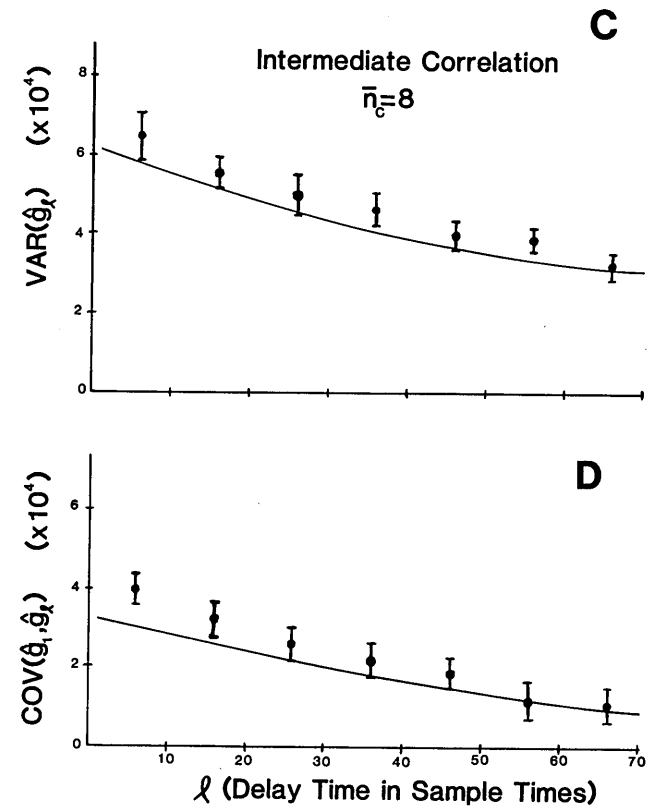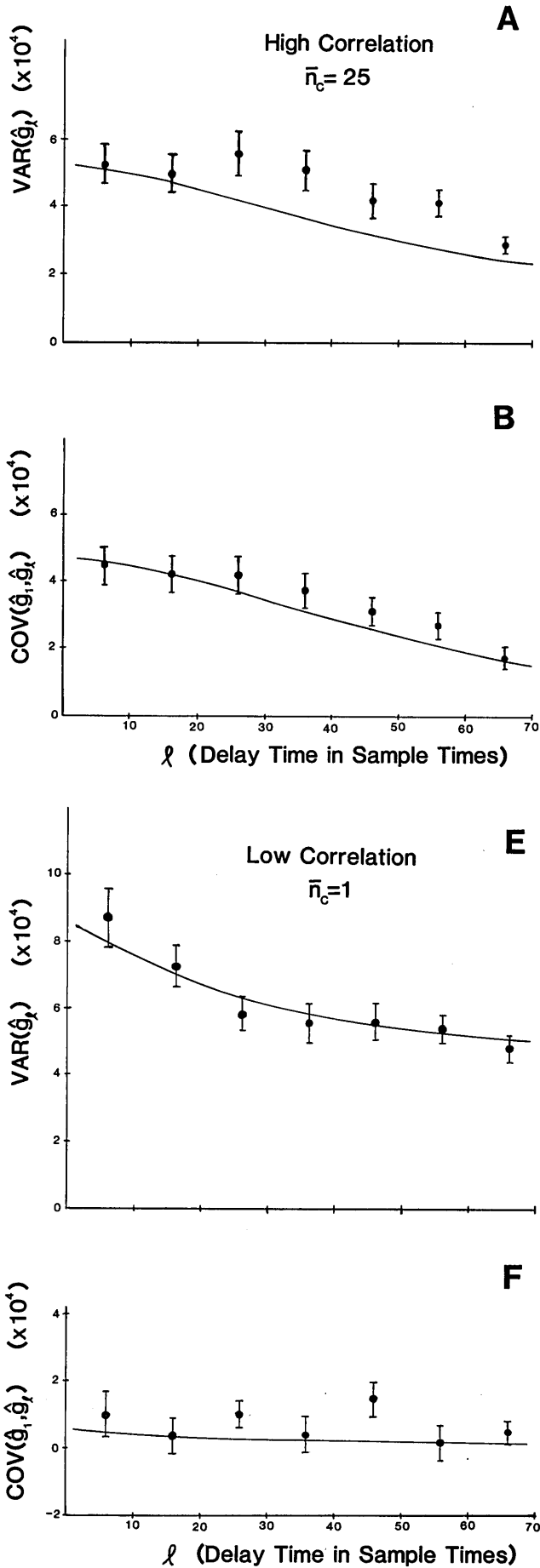$$\times \sum_{m=0}^{\infty}[(2m + 3)!!(2m + 2)(2m + 4)/6][\mathrm{Var}(\hat{n})/\overline{n}^2]^m.$$

$$(16)$$

Fig. 2.   Measured values of Var($\hat{g}_l$) and Cov($\hat{g}_k, \hat{g}_l$) are plotted and compared with the theoretical prediction (solid line) of Eq. (19). A He–Cd laser beam (441 nm) was scattered at 90° from 0.1-μm-diameter spheres in water at 25.4°C. The sample time T was 5 μsec, and the coherence time was 315 μsec. Error bars represent ±1 standard deviation. A, B, The average number of counts per coherence time $\bar{n}_c$ is 25, resulting in high correlation between channels of $g^{(2)}(\tau)$. C and D, E and F, The count rate was successively reduced so that $\bar{n}_c = 8$ and $\bar{n}_c = 1$, respectively, with accompanying reductions in the correlation between channels.

Keeping terms that are second order or lower in deviations of estimators from their mean values, Eq. (16) reduces to

$$\langle \hat{g}_k \hat{g}_l \rangle = (\bar{n})^{-4} \text{Cov}(\hat{G}_k, \hat{G}_l) + g_k g_l + 10 g_k g_l \text{Var}(\hat{n})/\bar{n}^2$$
$$- 4(\bar{n})^{-3}[g_k \text{Cov}(\hat{G}_l, \hat{n}) + \text{Cov}(\hat{G}_k, \hat{n}) g_l]. \quad (17)$$

Continuing to neglect terms of fourth order or higher in deviations of estimators from their means, and using the expression in Eq. (11) for BIAS$_l$, we have finally

$$\text{Cov}(\hat{g}_k, \hat{g}_l) = \langle \hat{g}_k \hat{g}_l \rangle - \langle \hat{g}_k \rangle \langle \hat{g}_l \rangle$$
$$= \langle \hat{g}_k \hat{g}_l \rangle - g_k g_l - g_k \text{BIAS}_l - \text{BIAS}_k g_l$$
$$= (\bar{n})^{-4} \text{Cov}(\hat{G}_k, \hat{G}_l) + 4 g_k g_l \text{Var}(\hat{n})/\bar{n}^2$$
$$- 2(\bar{n})^{-3} [g_k \text{Cov}(\hat{G}_l, \hat{n}) + \text{Cov}(\hat{G}_k, \hat{n}) g_l].$$
$$(18)$$

Equation (18) agrees with Eqs. (9)–(11) of Ref. 9.   When $k = l$, Eq. (18) reduces to Eq. (10) of Ref. 10.   In the case of a Lorentzian spectrum and for $N_c \gg 1 \gg \gamma = T/\tau_c$, we find for $k \leq l$ that

$$\mathrm{Cov}(\hat{g}_k, \hat{g}_l) = N_c^{-1}((1/\bar{n}\bar{n}_c)[1 + a_2 \exp(-2\gamma l)]\delta_{kl} + (2/\bar{n}_c)[a_2 + 2a_3 \exp(-2\gamma l)]\delta_{kl}$$
$$+ (2/\bar{n}_c)\{a_2 \exp[-2\gamma(l-k)] + 2a_3 \exp(-2\gamma l)\}(1 - \delta_{kl})$$
$$+ (2/\bar{n}_c)(a_2 - 2a_2^2 + 2a_3) \exp[-2\gamma(l+k)] + a_2^2[\gamma(l-k) + \tfrac{1}{2}] \; \exp\{-2\gamma(l-k)$$
$$+ 4a_4[\gamma(l-k) + 1]\exp(-2\gamma l) + (a_2^2 + 2a_4)[\gamma(l+k) + \tfrac{1}{2}]\exp(-2\gamma(l+k)\}$$
$$+ 4\{a_2{}^3 - a_2 a_3[\gamma(l+k) + 2]\}\exp[-2\gamma(l+k)]). \tag{19}$$

For a point detector the a's $= 1$, and Eq. (19) reduces to Eq. (23) of Ref. 9. Notice that for $\bar{n} \ll \bar{n}_c \lesssim 1$ the first term of Eq. (19) dominates, and the channels of $g^{(2)}(\tau)$ are uncorrelated. However, when $\bar{n}_c \gg \bar{n} \gtrsim 1$ the off-diagonal terms of the covariance matrix are comparable to the diagonal terms, and the channels of $g^{(2)}(\tau)$ are highly correlated. Also note that when $\bar{n}_c \gg \bar{n} \gtrsim 1$ an increase in the photocount rate (increase in $\bar{n}_c$ and $\bar{n}$) does little to improve the statistical accuracy of $g^{(2)}(\tau)$; only an increase in the correlator run duration (increase in N_c) enhances the accuracy.

In order to compare the theoretical predictions of Eq. (19) with measured values, we scattered the beam of a He–Cd laser ($\lambda = 441$ nm) from 0.1-μm spheres in water. The results are presented in Fig. 2. Note that the decrease in $\mathrm{Cov}(\hat{g}_k, \hat{g}_l)$ as $\bar{n}_c$ is reduced from 25 to 1. The agreement between the theoretical predictions and measured values is fair, although the measured values tend to exceed the predictions slightly. Extraneous effects such as laser-output fluctuations have not been included in the theoretical calculations.

4. PROCEDURE FOR THE ANALYSIS OF CORRELATED DATA

A typical PCS data-analysis procedure employs a least-squares fit of $g^{(2)}(\tau)$ to some expected functional form $f(\tau)$. Depending on the choice of $f(\tau)$, a linear or a nonlinear technique will be used to minimize χ^2:

$$\chi^2 = v^\dagger M^{-1} v, \tag{20}$$

where v is the column vector

$$v = \begin{bmatrix} (g_1 - f_1) \\ \vdots \\ (g_l - f_l) \\ \vdots \\ (g_L - f_L) \end{bmatrix}$$

and M is the $L \times L$ covariance matrix $M_{kl} = \langle(\hat{g}_k - \langle\hat{g}_k\rangle)(\hat{g}_l - \hat{g}_l)\rangle$. If the channels of $g^{(2)}(\tau)$ are uncorrelated, then the covariance matrix M is diagonal, and χ^2 reduces to

$$\chi^2 = \sum_{l=1}^{L} (g_l - f_l)^2/\sigma_l^2, \tag{21}$$

where the σ_l^2 are the variances M_{ll} of the channels of $g^{(2)}(\tau)$. PCS practitioners typically perform a number of correlator runs and calculate sample variances for the σ_l^2. If these measured values of σ_l^2 are used to calculate χ^2 according to Eq. (21), and if the channels are unexpectedly (positively) correlated, then clearly the minimum value of χ^2/DF will be erroneously low since many significant (positive) terms in Eq. (20) have been omitted. A simple way to correct this procedure is to use measured values not only for the variances $\sigma_l^2 = M_{ll}$ but also for the covariances M_{kl} ($k \neq l$). The matrix M is then inverted and used in the complete expression for χ^2, Eq. (20), and the correct minimum value for χ^2/DF is then

obtained. An alternative procedure would be to use the theoretical predictions of Eq. (19) for M_{kl}; after all, we demonstrated in Section 3 that the predictions of Eq. (19) are reliable. We have tried both procedures and have found that the time required to measure the M_{kl} with sufficient statistical accuracy to use M^{-1} as weighting in Eq. (20) is prohibitive. In order to measure the spectral parameter τ_c^{-1} to some desired accuracy, the measurement time using theoretical values for M_{kl} is far less (roughly a factor of 10) than the time required to employ measured values. We therefore suggest the following outline for a data analysis procedure: (1) Use Eq. (19) to evaluate the covariance matrix, (2) invert the covariance matrix, and (3) use a linear- or a nonlinear-regression technique to minimize χ^2 given by Eq. (20).

We shall now examine the performance of a data-analysis procedure that utilizes the theoretical expression for the M_{kl}. The data to be analyzed consist of 10 measurements (correlator runs) of $g^{(2)}(\tau)$ at each of three photocount rates corresponding to $\bar{n}_c = 1, 8, 25$. The beam from a He–Cd laser ($\lambda = 441$ nm) was scattered from 0.109 μm-diameter spheres in water at 25.4°C, and the photocount rate of light scattered at an angle of 90° was fed into a 128-channel 4-bit correlator (Langley-Ford Instruments Model LFI-128). In all measurements reported in this paper, the correlator yielded the full (not scaled) photocount-correlation function. Data from the correlator were transferred to a Digital Equipment Corporation LSI-11 computer for analysis. The sample time T was 5 μsec, and the coherence time was approximately 315 μsec. Ten points from each correlator run ($\tau = lT$; $l = 1, 8, 15, 22, \ldots, 64$) were chosen for the fitting procedure. The .fitting function was

$$f_l = 1 + c_0 \exp[-2(c_1 l + c_2 l^2 + c_3 l^3)], \tag{22}$$

where c_0 is the amplitude of $g^{(2)}(\tau)$ and is determined in these measurements primarily by the detector area ($c_0 = 0.84$), c_1 is τ_c^{-1} in units of inverse sample times, and the significance of c_2 and c_3 are discussed by Koppel.[8] The elements of the 10×10 covariance matrix M were calculated using Eq. (19), the matrix was inverted, and χ^2 was calculated using Eq. (20). A common nonlinear-regression technique described by Bevington[13] (program CURFIT) was used to minimize χ^2.

The results of the data analysis at each value of $\bar{n}_c$ are presented in Table 1. For each value of $\bar{n}_c$ three methods of analysis were used. First, the presumably "correct" procedure described in the previous paragraph was used to fit each of the 10 correlator runs, and the resulting values for χ^2/DF and $\tau_c^{-1}(c_1)$ were averaged and are presented in Table 1. The uncertainty listed for τ_c^{-1} is the sample standard deviation of the mean for the 10 correlator runs. An F test[13] was used to determine whether the inclusion of each parameter in the fitting function, Eq. (22), was justified. The term "linear" in Table 1 means that only c_0 and the coefficient c_1 of the linear term in the exponent of Eq. (22) were kept. Values for the average number of photocounts recorded in a coherence

Table 1.   Data-Analysis Results

$\bar{n}_c$	N_c	Method	χ^2/DF	F Test	τ_c^{-1} (sec^{-1})
25	3.2×10^3	correct	1.17	10 of 10 linear	3176 ± 53 (1.7%)
25	3.2×10^3	naive	0.20	7 of 10 linear	3120 ± 57 (1.8%)
25	3.2×10^3	mixed	1.20	10 of 10 linear	3112 ± 43 (1.4%)
8	6.3×10^3	correct	1.07	10 of 10 linear	3140 ± 53 (1.7%)
8	6.3×10^3	naive	0.59	10 of 10 linear	3132 ± 49 (1.6%)
8	6.3×10^3	mixed	1.15	10 of 10 linear	3126 ± 42 (1.3%)
1	1.6×10^5	correct	0.86	10 of 10 linear	3172 ± 35 (1.1%)
1	1.6×10^5	naive	0.83	10 of 10 linear	3172 ± 35 (1.1%)
1	1.6×10^5	mixed	0.89	10 of 10 linear	3172 ± 40 (1.3%)

time ($\bar{n}_c$) and the number of coherence times in each correlator run duration (N_c) are also listed in Table 1.

A second procedure was used for data analysis and is labeled "naive" in Table 1. In this procedure only the variances σ_l^2 were evaluated using Eq. (19), and χ^2 was calculated according to Eq. (21). It is not surprising to see correlated data ($\bar{n}_c = 25$) yield an erroneously low value for χ^2/DF. Consequently, the F test results suggest that 3 of the 10 fits should keep the coefficient c_2 of the quadratic term in the exponent of Eq. (22). The results of the naive and correct procedures are nearly identical for uncorrelated data ($\bar{n}_c = 1$).

A third procedure used is labeled "mixed" in Table 1. For each value of $\bar{n}_c$ there are 10 correlator runs, and from each run 10 channels were selected for fitting. From these 100 correlation function channels it is possible to form 10 mixed correlation functions, which have one channel from each of the 10 original correlator runs. When $\bar{n}_c = 25$ the 10 original correlator runs represent 10 statistically independent measurements of $g^{(2)}(\tau)$, each measurement consisting of 10 correlated channels. However, each mixed correlation function consists of 10 uncorrelated channels, although different mixed correlation functions are not statistically independent. In the "mixed" data-analysis procedure we used the theoretical predictions for the variances $\sigma_l^2 = M_{ll}$ and calculated χ^2 according to Eq. (21) using the mixed uncorrelated channels for the g_l. This procedure should be valid, as is borne out by the results listed in Table 1. The only problem occurs in averaging the results of the 10 fits to the mixed correlation functions, since it is not clear how to form a sample variance for τ_c^{-1} from the 10 statistically dependent values. We have ignored this problem for the purposes of Table 1 and have formed the sample variance assuming the 10 values are uncorrelated. It is satisfying to see that the "correct" and "mixed" procedures yield similar results.

The expected value of τ_c^{-1} can be calculated using nominal values for the temperature and viscosity of the water and the diameter of the spheres. Using the room temperature of 25.4 ± 1.0°C, a corresponding water viscosity of 0.882 ± 0.020 centipoise, a refractive index for water of 1.3328, the sphere diameter of 0.1090 ± 0.0027 μm (Dow Chemical Company), and a scattering angle of 90°, a value for τ_c^{-1} is found of 3279 ± 75 sec^{-1}. This expected value for τ_c^{-1} is in reasonable agreement with the values listed in Table 1.

The observed uncertainties in the measured values of τ_c^{-1} can be compared with theoretical predictions. Jakeman *et al.*[10] estimate the uncertainty in τ_c^{-1} when the correlation function channels are uncorrelated [see their Eq. (36)]. Saleh and Cardoso[9] estimate the uncertainty in τ_c^{-1} in the case of

Table 2.   Comparison of Estimated and Observed Uncertainties in τ_c^{-1}

		$\delta(\tau_c^{-1})/(\tau_c^{-1})$ (%)		
$\bar{n}_c$	N_c	Observed	Estimated (Correlated)	Estimated (Uncorrelated)
25	3.2×10^3	1.7	1.4	1.1
8	6.3×10^3	1.7	1.4	1.2
1	1.6×10^5	1.1	1.4	1.4

correlated channels [see their Eq. (46)]. Following these two references, we have calculated the expected uncertainties in τ_c^{-1} for the three experimental situations treated in Table 1. We have used Eq. (19) to calculate the covariance of channels and have calculated the expected uncertainty in τ_c^{-1} corresponding to the sample standard deviation of the mean of 10 correlator runs. We compare the expected and observed uncertainties in Table 2. Note that, when channel correlation is appreciable, the estimated uncertainty is greater when correlation is taken into account and that these estimates agree somewhat better with the observed uncertainties then do the estimates that do not account for channel correlation.

5.   SUMMARY AND DISCUSSION

We have calculated the statistical bias and channel correlation of $g^{(2)}(\tau)$ for the case of a stationary, cross-spectrally pure, Gaussian scattered field with a Lorentzian spectrum. Our treatment is valid for a detector area of arbitrary extent but assumes that the correlator run duration is much longer than a coherence time that is much longer than a sample time.

Experimental results confirm our theoretical calculations, which predict that the bias of $g^{(2)}(\tau)$ becomes significant when the correlator run duration is less than approximately 200 coherence times. This result holds when the average counts per coherence time ($\bar{n}_c$) is greater than or equal to 1. When $\bar{n}_c \lesssim 1$ the bias is appreciable for even longer correlator run durations. Intuitively, the uncertainty in the average counts per sample time ($\bar{n}$) becomes sufficiently large for short correlator run durations that $g^{(2)}(\tau)$ (which is normalized by $\bar{n}^2$) becomes biased. The bias must be taken into account in order to deduce an undistorted spectrum for the scattered light.

Experimental results also confirm our theoretical calculations of the channel correlation of $g^{(2)}(\tau)$. We observe a transition from highly correlated channels when $\bar{n}_c = 25$ to uncorrelated channels when $\bar{n}_c = 1$. Channel correlation can be explained intuitively in the following way. When $\bar{n}_c$ is

large, a burst of photocounts (e.g., 25 to 100) lasting approximately a coherence time contributes to the values of many channels of $g^{(2)}(\tau)$ [see Eq. (1)]. Since this burst of photocounts corresponds to a single intensity fluctuation (a single statistical event), it is clear that the final values of the channels of $g^{(2)}(\tau)$ are formed from many of the same statistical events, and hence the channels are correlated. However, when $\bar{n}_c \ll 1$ a burst of photocounts rarely contains more than two counts, which can only contribute to the value of one channel of $g^{(2)}(\tau)$. Hence the values of different channels are determined by different statistical events, and the channels are uncorrelated.

In the calculation of channel correlation we have accounted for the integration of scattered light intensity over a detector area of arbitrary extent. As a result, the theoretical calculations predict measured channel correlation for common experimental situations with sufficient accuracy to enable predicted values to be used as weighting in a least-squares data-analysis procedure. The use of predicted values avoids time-consuming measurements of channel correlation. Our analysis of measured values of $g^{(2)}(\tau)$ has demonstrated that reliable values of channel correlation must be used in regression procedures in order to evaluate meaningfully the assumed fitting function and the extent of polydispersity of the scatterers.

APPENDIX A: THE SPATIAL-INTEGRATION FACTORS

The effect of a finite detector area on the photocounting statistics of a stationary, cross-spectrally pure, Gaussian field has been discussed by Cantrell and Fields[14] and reviewed by Saleh.[15] We shall simply relate the spatial-integration factors appearing in Equations (12)–(14) to the first four moments of the number of photocounts recorded in a sample time: $\langle n \rangle$, $\langle n^2 \rangle$, $\langle n^3 \rangle$, and $\langle n^4 \rangle$. These moments are easily measured. Many commercially available correlators measure the photocount-distribution function $P(n)$ from which the moments can be calculated. We measured the moments with a scaler interfaced to a computer. When the sample time is much less than the coherence time ($\gamma = T/\tau_c \ll 1$), we find that

$$a_2 = (\langle n^2 \rangle - \langle n \rangle)/\langle n \rangle^2 - 1,$$

$$a_3 = (\tfrac{1}{2})[(\langle n^3 \rangle - 3\langle n^2 \rangle + 2\langle n \rangle)/\langle n \rangle^3 - 1 - 3a_2],$$

$$a_4 = (\tfrac{1}{6})[(\langle n^4 \rangle - 6\langle n^3 \rangle + 11\langle n^2 \rangle - 6\langle n \rangle)/\langle n \rangle^4$$
$$- 1 - 6a_2 - 3a_2{}^2 - 8a_3].$$

It should be noted that the spatial factors (a's) depend only on the size of the detector and the intensity distribution of the scattering volume. For a point detector $a_2 = a_3 = a_4 = 1$; for

a detector area of finite extent $0 < a$'s < 1. For the experimental arrangement employed in our measurements we found that $a_2 = 0.817$, $a_3 = 0.740$, and $a_4 = 0.645$.

ACKNOWLEDGMENTS

* Present address, P.O. Box 20047, New York, New York 10025.

REFERENCES

1. R. Pecora, "Doppler shifts in light scattering from pure liquids and polymer solutions," J. Chem. Phys. 40, 1604–1614 (1964).
2. H. Z. Cummins, N. Knable, and Y. Yeh, "Observation of diffusion broadening of Rayleigh scattered light," Phys. Rev. Lett. 12, 150–153 (1964).
3. R. C. Haskell and F. D. Carlson, "Quasi-elastic light scattering studies of single skeletal muscle fibers," Biophys. J. 33, 39–62 (1981).
4. J. Newman and F. D. Carlson, "Dynamic light scattering evidence for the flexibility of native muscle thin filaments," Biophys. J. 29, 37–48 (1980).
5. G. B. Stock, "The measurement of bacterial translation by photon correlation spectroscopy," Biophys. J. 22, 79–96 (1979).
6. J. Frost and H. Z. Cummins, "Motility assay of human sperm by photon correlation spectroscopy," Science 212, 1520–1522 (1981).
7. M. D. Stern, A. A. Kort, G. M. Bhatnagar, and E. G. Lakatta, "Scattered light intensity fluctuations in diastolic rat cardiac muscle caused by spontaneous Ca^{++} dependent cellular mechanical oscillations," J. Gen. Physiol. 82, 119–153 (1983)
8. D. E. Koppel, "Analysis of macromolecular polydispersity in intensity correlation spectroscopy: the method of cumulants," J. Chem. Phys. 57, 4814–4820 (1972).
9. B. E. A. Saleh and M. F. Cardoso, "The effect of channel correlation on the accuracy of photon counting digital autocorrelators," J. Phys. A 6, 1897–1909 (1973).
10. E. Jakeman, E. R. Pike, and S. Swain, "Statistical accuracy in the digital autocorrelation of photon counting fluctuations," J. Phys. A 4, 517–534 (1971).
11. A. J. Hughes, E. Jakeman, C. J. Oliver, and E. R. Pike, "Photon-correlation spectroscopy: dependence of linewidth error on normalization, clip level, detector area, sample time and count rate," J. Phys. A 6, 1327–1336 (1973).
12. C. J. Oliver, "The extraction of spectral parameters in photon-correlation spectroscopy," Adv. Phys. 27, 387–435 (1978).
13. P. R. Bevington, Data Reduction and Error Analysis for the Physical Sciences (McGraw-Hill, New York, 1969).
14. C. D. Cantrell and J. R. Fields, "Effect of spatial coherence on the photoelectric counting statistics of Gaussian light," Phys. Rev. A 7, 2063–2069 (1973).
15. B. Saleh, Photoelectron Statistics (Springer-Verlag, New York, 1978); see especially Sec. 5.2.2.