

Information and resolution in electromagnetic optical systems

Matthew R. Foreman* and Peter Török

Blackett Laboratory, Department of Physics, Imperial College London, Prince Consort Road, London, SW7 2BZ, United Kingdom

(Received 25 May 2010; published 25 October 2010)

Quantitative analysis can play a vital role in a number of polarization-based optical systems, yet to date no definition regarding resolution in the polarization domain exists. By adopting a stochastic framework, a suitable metric is developed in this article, allowing a number of polarimetric systems to be assessed and compared. In so doing, the performance dependencies of polarization-based systems are demonstrated and fundamental trends are identified.

DOI: [10.1103/PhysRevA.82.043835](https://doi.org/10.1103/PhysRevA.82.043835)

PACS number(s): 42.25.Ja, 89.70.Cf, 07.60.Fs

I. INTRODUCTION

Analysis of polarized light, or polarimetry, is a popular and useful tool in science today. Applications vary from astronomy, microscopy, remote sensing, and biomedical diagnosis [1–3] to more fundamental crystallographic, material, and single-molecule studies [4,5]. Polarization can also be utilized in quantum cryptography and communication [6]. Quantitative analysis of the polarization state of light output from a physical system of interest (e.g., light scattered from a biological sample), plays a key role in such studies; for example, in determining the depolarization or phase retardation [7] of a sample. Unfortunately, all practical systems are subject to noise, thus restricting the accuracy with which parameters of interest can be determined or the reliability of conclusions drawn. For example, searches for polarization signatures of quantum gravity in the cosmic background require high-precision polarimetric measurements [8], whilst error rates in wireless communications and polarimetric classification can be compromised by poor measurement accuracy [9,10]. Despite the growing prominence of polarization-based systems, no definition of resolution in a polarization domain exists in the literature, hence system capabilities cannot be easily assessed and compared. This article therefore principally aims to formulate a suitable definition of polarization resolution and to evaluate the performance of a number of common polarimetric architectures frequently found in the literature.

Resolution, by any definition, aims to encapsulate the limits to which something can be measured or, in other words, characterizes a region of uncertainty in measurement space. Consider, for example, Rayleigh's resolution criterion, which quantifies the minimum angular separation between which two point objects, imaged by an incoherent diffraction-limited imaging system, appear separated. Specifically, Rayleigh's criterion states that this minimum separation occurs when the first minimum in the image of a point source (e.g., an Airy pattern) coincides with the maximum of the image of a second point object. Theoretically, in a noiseless system two point objects can be infinitely resolved; for example, by deconvolving the image with the point spread function of the system. However, Rayleigh's criterion implicitly (although arbitrarily) makes an assumption of the minimum change of intensity that can be measured, and hence the extent to

which the central depression in the total intensity profile can be distinguished.

Similar principles can be applied to polarimetry; however, the appropriate region of uncertainty exists not in a spatial domain, but instead in the polarization domain. A convenient representation of the polarimetric measurement space and process is therefore required and is hence considered in Secs. II and III. Rayleigh's criterion is, however, fundamentally deterministic in nature and hence does not characterize the stochastic origin of uncertainty in any measurement. A stochastic framework founded in statistical estimation theory is therefore more appropriate, and is thus presented in Sec. IV and from which a definition of polarization resolution ultimately follows. Further alternative, but insightful, metrics are also constructed. Sections V and VI finally proceed to apply the developed theory to three polarimetric architectures commonly found in the literature. So doing highlights the contrasting characteristics of different measurement configurations, whilst simultaneously identifying important performance dependencies on source properties.

II. POINCARÉ AND STOKES SPACE

Polarized light (assumed henceforth to be quasimonochromatic and plane wave in nature) is frequently described using the Stokes formalism, whereby a Stokes vector

$$\mathbf{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} \langle |E_x|^2 + |E_y|^2 \rangle \\ \langle |E_x|^2 - |E_y|^2 \rangle \\ 2\langle \text{Re}[E_x E_y^*] \rangle \\ 2\langle \text{Im}[E_x E_y^*] \rangle \end{pmatrix} = S_0 \begin{pmatrix} 1 \\ P \cos 2\varepsilon \cos 2\vartheta \\ P \cos 2\varepsilon \sin 2\vartheta \\ P \sin 2\varepsilon \end{pmatrix}$$

defines a position with polar coordinates $(P, \pi/2 - 2\varepsilon, 2\vartheta)$ within a unit sphere known as the Poincaré sphere [11]. The individual Stokes parameters S_i ($i = 0, 1, 2, 3$) describe the total intensity of a light beam, the difference in the intensity of horizontal (E_x) and vertical (E_y) polarized components, the difference in intensity of components polarized at $\pm 45^\circ$ and, finally, the difference in intensity of right- and left-handed circularly polarized components, respectively. Angular braces $\langle \dots \rangle$ denote temporal averaging over a time interval τ , assumed long enough to ensure that the time average is independent of τ . P is known as the degree of polarization and is defined as the ratio of the intensity of the polarized component of a beam to the total intensity [i.e., $P = (S_1^2 + S_2^2 + S_3^2)^{1/2}/S_0$]

*matthew.foreman@imperial.ac.uk

($0 \leq P \leq 1$)]. Specifically, for fully polarized light ($P = 1$), the Stokes parameters define a position on the surface of the Poincaré sphere, whereby linearly polarized states are located on the equator, whilst points on the north (south) pole describe right- (left-) circularly polarized light. Partially polarized states of light ($P < 1$) are described by points lying within the sphere, with unpolarized light ultimately lying at the center of the sphere.

The Poincaré sphere exists in a Hilbert space \mathcal{P} , with coordinate axes $\{\check{S}_1, \check{S}_2, \check{S}_3\}$, where $\check{S}_i = S_i/S_0$ denotes the normalized Stokes parameters, which shall be referred to as Poincaré space, [see Fig. 1(a)]. Note that the caron notation

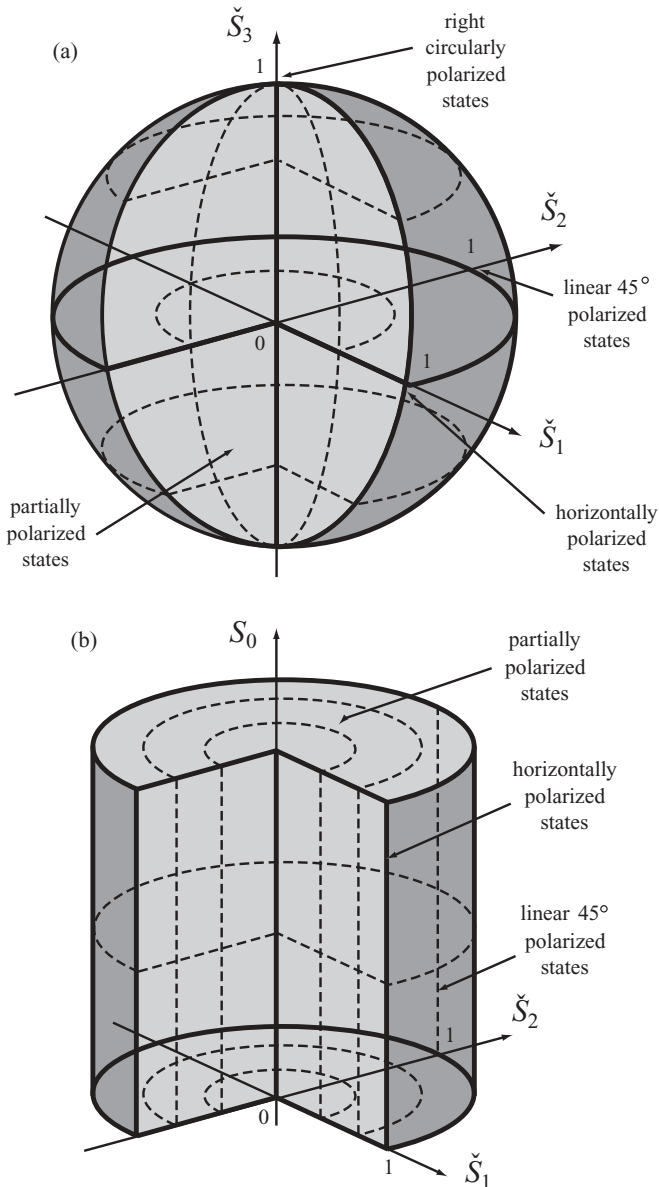


FIG. 1. Schematic of the (a) Poincaré-sphere and (b) Stokes-cylinder construction illustrated for linearly polarized light ($\check{S}_3 = 0$). The axis of the cylinder defines the intensity axis S_0 . Totally polarized states lie on the outer surface of both the Poincaré sphere and the Stokes cylinder (dark shading), with partially polarized states lying inside (lighter shading).

($\check{\cdot}$) has been used in contrast to the conventional hat ($\hat{\cdot}$) notation so as to avoid confusion with that used for statistical estimators in subsequent sections. Difficulties can, however, arise when considering the Poincaré-sphere representation since all four degrees of freedom in \mathbf{S} are not suitably depicted. When describing general partially polarized states, it is perhaps more appropriate to consider the Stokes vector as defining a position in a Hilbert space \mathcal{S} with coordinate axes $\{S_0, \check{S}_1, \check{S}_2, \check{S}_3\}$, which shall be referred to as Stokes space. Physically allowable Stokes vectors must satisfy the inequality $S_0^2 \geq S_1^2 + S_2^2 + S_3^2$ and, therefore, define a Stokes cylinder in Stokes space of unit radius, with the intensity S_0 defining the axis of the cylinder [see Fig. 1(b)]. A similar idea was proposed by Tyo [12] in which the coordinate axes were instead defined by $\{S_0, S_1, S_2, S_3\}$, meaning admissible Stokes vectors spanned a Stokes cone; however, use of normalized coordinates dissociates the polarization properties and intensity of a beam allowing greater insight to be gained into performance dependencies of polarimetric systems. A slice in Stokes space perpendicular to the S_0 axis yields the Poincaré sphere, whilst a cross section taken perpendicular to the \check{S}_3 axis (and through the origin) produces a cylinder with unit radius akin to that shown in Fig. 1(b). Points on the surface of this hypercylinder describe totally linearly polarized light, whilst points toward the central axis describe partially polarized states. Similar cross sections taken perpendicular to \check{S}_1 and \check{S}_2 also give hypercylinders.

III. POLARIMETRIC ANALYSIS

When polarized light is transmitted through a polarization-state analyzer (PSA) the transmitted intensity D can be found by projecting the input Stokes vector onto a measurement vector $\mathbf{T} = (1, T_1, T_2, T_3)^T / 2$ ($\mathbf{T} \in \mathcal{S}$), as defined by the analyzer configuration, whereby $D = \mathbf{T}^T \cdot \mathbf{S}$. Here, \mathbf{T} is normalized such that $T_1^2 + T_2^2 + T_3^2 = 1$, so that the transmitted intensity $0 \leq D \leq S_0$; that is, to ensure the analyzer is passive.

With prior knowledge of the analyzing state, as can be deduced from the polarization elements present in the PSA, it would theoretically be possible to estimate the Stokes vector of the incident light. In general however, multiple measurements are made to improve accuracy and remove ambiguities that may exist. For example, a division of amplitude polarimeter (DOAP), as originally proposed by Azzam [13], uses at least four different analyzers simultaneously, whilst a null ellipsometer uses an analyzer which is varied between sequential measurements [11]. Arranging the N_D respective transmitted intensities into an intensity vector $\mathbf{D} = (D_1, D_2, \dots, D_{N_D})$, the series of measurements can be described by the matrix equation

$$\mathbf{D} = \mathbb{V} \mathbf{T} \mathbf{S}, \quad (1)$$

where \mathbb{T} is a $N_D \times 4$ instrument matrix with rows corresponding to the N_D measurement states. To ensure that Eq. (1) can describe both sequential or simultaneous measurements, the diagonal matrix \mathbb{V} has been introduced to account for the beam splitting required for simultaneous measurements. Conservation of energy (assuming ideal optical elements) dictates that, for multiple simultaneous measurements, $\text{tr}[\mathbb{V}] = 1$. Alternatively, if \mathbf{D} is formed from N_D sequential

measurements, then $\mathbb{V} = \mathbb{I}$ where \mathbb{I} is the $N_D \times N_D$ identity matrix. Realistically, energy will be lost during propagation through an optical system from absorption and scattering, for example. Consequently $\text{tr}[\mathbb{V}]$ can then be used as a measure of the light efficiency of the PSA.

IV. MEASUREMENT UNCERTAINTY AND POLARIZATION RESOLUTION

A. Fisher information and the Cramér-Rao lower bound

Practically, the empirical intensities \mathbf{D} will possess a random component arising from noise present in the system (see Sec. V A for a discussion of some potential noise models), such that $\mathbf{D} = \mathbf{D}_0 + \Delta\mathbf{D}$ is a random variable ($\Delta\mathbf{D}$ will be assumed to be a zero-mean random variable throughout this work). Consequently, it may not be possible to solve Eq. (1) for \mathbf{S} exactly, but instead a “best guess” $\hat{\mathbf{S}}$ must be formed for which a multitude of alternative estimation strategies exist. For example, the most immediately apparent solution to Eq. (1) is given by $\hat{\mathbf{S}} = [\mathbb{V}\mathbb{T}]^+\mathbf{D}$, where $^+$ denotes the Moore-Penrose matrix pseudoinverse. Under a Gaussian noise regime the pseudoinverse solution minimizes the mean square error between the true value of \mathbf{S} and the different realizations of $\hat{\mathbf{S}}$ (which inherit stochasticity from the random nature of \mathbf{D}).

More generally, the quality of any particular estimator $\hat{\mathbf{w}}$ of a set of parameters \mathbf{w} can be assessed by considering its covariance matrix $\mathbb{K}_{\mathbf{w}}$. The on-diagonal terms of the covariance matrix describe the spread of estimates of each parameter w_i , where a smaller spread implies a more accurate estimator, whilst the off-diagonal terms quantify the dependence of the estimate of one parameter, say w_i , on another w_j ($i \neq j$). Pivotal in the theory of statistical estimation was the formulation of a lower bound on the covariance matrix of any estimator, known as the Cramér-Rao lower bound (CRLB) [14,15]. The CRLB, in its most commonly used form, is valid for unbiased estimators whereby it is possible to say that the variance of any estimate made of the parameter vector \mathbf{w} is bounded by

$$\mathbb{K}_{\mathbf{w}} \geq \mathbb{J}_{\mathbf{w}}^{-1}, \quad (2)$$

where the inequality implies the difference of the two matrices is positive definite and does not necessarily hold element-wise. $\mathbb{J}_{\mathbf{w}}$ is a $N_w \times N_w$ matrix known as the Fisher information matrix (FIM) [16,17] and defined by

$$\mathbb{J}_{\mathbf{w}} = E \left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x}|\mathbf{w})}{\partial \mathbf{w}} \right)^T \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x}|\mathbf{w})}{\partial \mathbf{w}} \right], \quad (3)$$

where $f_{\mathbf{X}}(\mathbf{x}|\mathbf{w})$ is the probability density function (PDF) describing the likelihood that the random variable $\mathbf{X} = \mathbf{x}$, as parameterized by \mathbf{w} , and $E[\cdot \cdot \cdot]$ denotes the ensemble average over all possible realizations of \mathbf{x} . It should also be noted that the matrix calculus conventions expounded in [18] are used in this work.

Since Eq. (3) depends only on the type of noise present in a system through the PDF $f_{\mathbf{X}}(\mathbf{x}|\mathbf{w})$, the CRLB can be considered a fundamental statistical limit on the quality of any measurement system. An estimator which achieves the CRLB; that is, for which $\mathbb{K}_{\mathbf{w}} = \mathbb{J}_{\mathbf{w}}^{-1}$, is called efficient. However, the question arises as to the existence of such an estimator.

A well-documented result in statistical fields is that if an efficient estimator exists then it is the maximum-likelihood estimator (see, e.g., [19]); however, if no such estimator exists then the maximum-likelihood estimator is both asymptotically unbiased and asymptotically efficient as the number of data points taken tends to infinity. These desirable properties are not mirrored by other estimators.

The equations given thus far are valid only for a particular value of \mathbf{w} ; however, the parameter values may differ between different experimental setups or measurements. Adopting a Bayesian viewpoint, the parameter vector \mathbf{w} is then considered to be a random variable with an associated PDF $f_{\mathbf{w}}(\mathbf{w})$ (known as a prior PDF). Accordingly, it is possible to modify the definition of the FIM to accommodate this random behavior such that the Bayesian FIM is given by

$$\tilde{\mathbb{J}}_{\mathbf{w}} = E_{\mathbf{w}}[\mathbb{J}_{\mathbf{w}}] + \mathbb{J}_{\mathbf{w}}^p, \quad (4)$$

where $\mathbb{J}_{\mathbf{w}}$ is the deterministic FIM given by Eq. (3) and $\mathbb{J}_{\mathbf{w}}^p$ depends only on the prior PDF $f_{\mathbf{w}}(\mathbf{w})$ via

$$\mathbb{J}_{\mathbf{w}}^p = E_{\mathbf{w}} \left[\left(\frac{\partial}{\partial \mathbf{w}} \ln f_{\mathbf{w}}(\mathbf{w}) \right)^T \frac{\partial}{\partial \mathbf{w}} \ln f_{\mathbf{w}}(\mathbf{w}) \right]. \quad (5)$$

A Bayesian CRLB can thus be shown to hold, as was originally done by van Trees [20], whereby

$$\mathbb{K}_{\mathbf{w}} \geq \tilde{\mathbb{J}}_{\mathbf{w}}^{-1}, \quad (6)$$

(the tilde notation will be used throughout to represent Bayesian metrics). If no *a priori* knowledge about the possible values of the random parameter is possessed, as is likely in many experimental scenarios, it can be argued that the prior PDF should be nearly flat, such that any estimator formed will not cluster around any particular value. In the limit a uniform PDF over the admissible values of \mathbf{w} can be used. Such a PDF is known as a noninformative PDF. The assumption of maximal ignorance can be relaxed, as discussed in [21], but this case will not be considered in this work.

Again, the question of the existence of efficient estimators arises in the Bayesian paradigm. Fortunately, it can be shown [20] that the maximum *a posteriori* estimator is, in many respects, the Bayesian equivalent of the maximum-likelihood estimator. As such, if an efficient estimator exists in the Bayesian sense then it will be the maximum *a posteriori* estimator, otherwise the maximum *a posteriori* estimator is asymptotically efficient and hence the Bayesian CRLB is achievable.

B. Resolution in Stokes space

As a metric of performance, a matrix quantity, such as the FIM or estimator covariance matrix, is not ideal since its interpretation is often nontrivial. Scalar measures are thus preferable, therefore prompting the introduction of suitable quantities. Since the parameter estimate $\hat{\mathbf{w}}$ is derived from random experimental data, $\hat{\mathbf{w}}$ is itself a random variable. Each experimental realization, perturbed by differing noise, hence defines a different point in the N_w -dimensional Hilbert space in which \mathbf{w} lies. For example, if attempting to estimate the Stokes vector of polarized light from noisy intensity measurements, whereupon $\mathbf{w} = \mathbf{S}$, each estimate $\hat{\mathbf{S}}$ defines

a position in Stokes space. If the estimator is unbiased, the random distribution of estimates will be centered on the true parameter value. The eigensystem of the FIM then defines the axes of a set of concentric “ellipsoids of minimum uncertainty” (in the sense of the CRLB) in Hilbert space, defined by $(\hat{\mathbf{w}} - \mathbf{w})^\dagger \mathbb{J}(\hat{\mathbf{w}} - \mathbf{w}) = c^2$. The parameter c dictates the fraction of the estimates resulting from repeated experiments which lie within the ellipsoid [19]. For example, if an unbiased efficient estimator $\hat{\mathbf{w}}$ with covariance matrix \mathbb{J}^{-1} was normally distributed, then the probability that a particular estimate would lie within the region $c^2 \leq c_0^2$ could be found by integrating the N_w -dimensional χ^2 -squared probability distribution from 0 to c^2 .

With these considerations in mind, it is apparent that the volume of the so-called ellipsoids of concentration, given by

$$V_{\min} = V_{N_w} \sqrt{c^{N_w} |\mathbb{J}_{\mathbf{w}}^{-1}|} = V_{N_w} \sqrt{\frac{c^{N_w}}{|\mathbb{J}_{\mathbf{w}}|}}, \quad (7)$$

where V_{N_w} is the volume of the N_w -dimensional unit hypersphere, describe a region of uncertainty in the relevant Hilbert space. Letting $\mathbf{w} = (\check{S}_1, \check{S}_2, \check{S}_3, S_0)$ (the reordering is for later convenience), whereby $N_w = 4$ and $V_4 = \pi^2/2$, the region of uncertainty in Stokes space has a volume

$$V_{\mathbf{w}} = \frac{\pi^2}{2} \sqrt{\frac{c^4}{|\mathbb{J}_{\mathbf{w}}|}}, \quad (8)$$

which hence defines a polarization resolution in that space. A suitable choice of c in this definition is no less arbitrary than, for example, the Rayleigh criterion, as it is merely a measure of what is acceptable to the end user. A value of 0.9 will henceforth be assumed.

Typically, $\mathbb{J}_{\mathbf{w}}$ will be dependent on the state of polarization being measured, hence motivating the further definition of a Bayesian polarization resolution $\tilde{V}_{\mathbf{w}}$ defined analogously to Eq. (8), in which the Bayesian FIM $\mathbb{J}_{\mathbf{w}} = E_{\mathbf{w}}[\mathbb{J}_{\mathbf{w}}]$ is used.

In polarimetry, however, the absolute intensity of the light may be of secondary or little importance, since the state of polarization is fully specified by the three variables $\{\check{s}_1, \check{s}_2, \check{s}_3\}$. The estimation problem thus reduces to inferring a position in the 3-dimensional Hilbert space spanned by the parameter vector $\mathbf{u} = (\check{s}_1, \check{s}_2, \check{s}_3)$ corresponding to Poincaré space. Unfortunately, this process still necessitates the estimation of the total intensity of the beam s_0 , be it implicit or explicit. Ultimately, the requisite estimation of s_0 reduces the polarization accuracy obtainable in Poincaré space, the extent of which can be assessed by partitioning the FIM; namely,

$$\mathbb{J}_{\mathbf{w}} = \begin{pmatrix} \mathbb{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{12}^T & J_{22} \end{pmatrix}, \quad (9)$$

where \mathbb{J}_{11} is a 3×3 reduced FIM, \mathbf{J}_{12} is a 3×1 column vector describing the cross correlations between estimates of \mathbf{u} and s_0 , and J_{22} is a scalar whose reciprocal describes the accuracy achievable for any estimate of s_0 via the CRLB. For any single state of polarization $V_{\mathbf{w}}/V_{\mathbf{u}} = (3\pi/8)\sqrt{c/|J_{22}|}$, as follows from $|\mathbb{J}_{\mathbf{u}}| = |\mathbf{J}_{\mathbf{w}}|/|J_{22}|$. The Bayesian FIM relevant to the estimation of \mathbf{u} is then given by [19]

$$\tilde{\mathbb{J}}_{\mathbf{u}} = E_{\mathbf{u}|s_0}[\mathbb{J}_{\mathbf{u}}] = E_{\mathbf{u}|s_0}[\mathbb{J}_{11} - \mathbf{J}_{12}^T J_{22}^{-1} \mathbf{J}_{12}], \quad (10)$$

where $E_{\mathbf{u}|s_0}[\dots]$ denotes averaging with respect to \mathbf{u} for a given total intensity. The cross correlations between estimates of \mathbf{u} and s_0 are thus seen to cause a reduction in the polarization resolution. The reduced Bayesian polarization resolution in Poincaré space is subsequently given by

$$\tilde{V}_{\mathbf{u}} = \frac{4\pi}{3} \sqrt{\frac{c^3}{|\tilde{\mathbb{J}}_{\mathbf{u}}|}}. \quad (11)$$

By partitioning $\mathbb{J}_{\mathbf{w}}$ in different ways, the treatment can be extended to situations in which not all polarization parameters are desired; for example, if considering only linear polarimeters.

C. Polarization encoding and degrees of freedom

Multiplexing of an optical signal, whereby information is encoded using different degrees of freedom of light, provides a means to increase information storage and transmission rates in optical systems. For example, different wavelengths can be used to send multiple signals along optical fibers [22] in so-called wavelength-division multiplexing. Fundamentally, for wavelength-division multiplexing, the number of different wavelengths (or, more generally, the number of channels or degrees of freedom) depends on the bandwidth of the channel and the extent of interchannel interference (crosstalk) that can be tolerated. For example, in fiber optic telecommunication networks which operate in the 1480- to 1600-nm low-loss window of silica glass, the international recommendation is for a wavelength spacing of 0.8 nm ranging from 1537 to 1563 nm so as to give 32 channels with acceptable levels of crosstalk [23].

Polarization encoding is also possible; however, it is almost exclusively considered in the context of only two orthogonal states of polarization (see, e.g., [24]). Such analysis is perhaps natural in the sense that crosstalk between the two degrees of freedom is zero in the ideal case, but it automatically forsakes the possibilities afforded by encoding over the entirety of Poincaré space (or Stokes space if amplitude modulation is also employed). Given the ability of polarimeters to distinguish multiple states of polarization, it is hence logical to investigate the number of degrees of freedom within polarization-based systems, as will be determined by the size of the polarization domain and the polarization resolution of the PSA. In this vein, the number of distinguishable states is defined here to be the ratio of the volume of uncertainty in polarization space before a measurement to the volume of uncertainty after a measurement (and hence can equally be called a metric of fractional accuracy). Explicitly, the number of degrees of freedom is then given by

$$\tilde{\mathcal{N}}_S = \tilde{\mathcal{A}}_w = \frac{V_S}{\tilde{V}_{\mathbf{w}}}, \quad (12)$$

and

$$\tilde{\mathcal{N}}_P = \tilde{\mathcal{A}}_u = \frac{V_P}{\tilde{V}_{\mathbf{u}}}, \quad (13)$$

when considering encoding in Stokes and (reduced) Poincaré space, respectively. An intuitive analog to this definition can be found in optical imaging, whereby the uncertainty in an object's position before a measurement is merely the field of view of the imaging system, whereas afterwards, assuming

a diffraction-limited system with circular aperture, it is the area of the Airy disk. Similarly, the uncertainty before a polarimetric measurement is the entirety of the associated Hilbert space. The volume is thus easily calculable using the Lebesgue measure and is given by $V_S = \frac{4\pi}{3}\mathcal{D}$ for Stokes space, where \mathcal{D} is the dynamic range of the photodetectors and $V_P = \frac{4\pi}{3}$ for Poincaré space. If one or more of the polarization parameters are known *a priori*, only the volume of space spanned by the unknown parameters need be considered.

Finally, a local accuracy can also be defined analogously to Eqs. (12) and (13) if the FIM before Bayesian averaging (i.e., \mathbb{J}_w or \mathbb{J}_u) is used to define the volume of the ellipsoid of concentration for a given incident polarization state.

D. Efficiency of observation

When performing experiments in conditions with limited light levels; for example, in single-molecule studies [25], it is important to utilize the detected photons as efficiently as possible. In the context of polarimetric experiments, this implies achieving the greatest accuracy, or polarization resolution, per photon. Physical limits, however, exist regarding the extent to which this can be achieved. To establish these limits, note that the reciprocal of the accuracy of a PSA can be considered as the fractional volume of uncertainty in polarization space, or the probability of measuring a state of polarization lying within the same volume, if all polarization states were equally likely. One can thus (following Shannon [26]) associate an information gain from a polarimetric observation as the logarithm of the accuracy; namely,

$$\tilde{\mathcal{I}} = \log_2 \tilde{\mathcal{A}} = -\log_2 \left(\frac{V_u}{\tilde{V}_p} \right). \quad (14)$$

The relationship between physical entropy in the thermodynamical sense and information has been known for many years and was first recognized by Szilard [27] and later applied by Brillouin [28,29]. The relationship states that information $\tilde{\mathcal{I}}$ about a system can only be obtained if there is an increase in entropy ΔH such that

$$\tilde{\mathcal{I}} \leq \frac{\Delta H}{k_B \ln 2}, \quad (15)$$

where $k_B = 1.381 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-1} \text{ K}^{-1}$ is Boltzmann's constant. Equality is only achieved for a reversible observation. From inequality (15), it is thus possible to define the efficiency of observation η ($0 \leq \eta \leq 1$):

$$\eta = \frac{\tilde{\mathcal{I}} k_B \ln 2}{\Delta H}. \quad (16)$$

Consider then a single optical detector which makes an observation by absorption of n_{i0} photons with mean energy $h\nu_0$, where $n_{i0} = E[n_i]$ is the mean number of photons absorbed by the i th detector.¹ The total energy absorbed will eventually be dissipated as heat, corresponding to an

¹Entropy is an average property of a system and hence it is sufficient to consider the average number of photons absorbed n_{i0} , as opposed to a particular realization of the observation process in which n_i photons are absorbed.

increase in the entropy of the detector. The second law of thermodynamics then dictates that $\Delta H_i = n_{i0}h\nu_0/\Theta$, where Θ is the thermal noise temperature (i.e., ambient temperature) of the detector. There will, however, be an entropy cost for each measurement made such that the total entropy cost is given by

$$\Delta H = \sum_{i=0}^{N_D-1} \Delta H_i + \Delta H_a, \quad (17)$$

where ΔH_a represents the entropy cost associated with photons that are not absorbed in the detectors. Ultimately, these "lost" photons will also be absorbed by some material body at temperature Θ_0 and again be dissipated as heat, such that $\Delta H_a = (1 - \text{tr}[\mathbb{V}])s_0/\Theta_0$ for simultaneous measurements or, alternatively, $\Delta H_a = (N_D - \text{tr}[\mathbb{V}])s_0/\Theta_0$ for sequential observations. If all photons are absorbed by photodetectors (i.e., the PSA is 100% light efficient), the efficiency of observation is given by

$$\eta = \frac{k_B \Theta \ln 2}{n_0 h \nu_0} \log_2 \left(\frac{V_P}{\tilde{V}_u} \right), \quad (18)$$

where $s_0 = \sum_i n_{i0}h\nu_0 = n_0h\nu_0$.

V. EXAMPLES

Accuracy, information, and efficiency of observation have all been shown to be dependent on the FIM (which is averaged under the assumption of maximal ignorance to form the Bayesian FIM) and, as such, all that remains to quantify system performance in polarization measurements is to calculate \mathbb{J}_w . However, this requires making some assertions regarding the type of noise present in the system and of the PSA configuration. In the following numerical calculations the reduced FIM (and Bayesian FIM) associated with the estimation of \mathbf{u} will only be considered so as to elucidate the polarization-dependent performance characteristics of different PSAs.

A. Noise models and Fisher information

In what follows, two noise regimes will be considered; namely, Poisson and Gaussian statistics. The first example discussed considers the quantization of classical light, which produces Poisson-distributed noise on the detector with variance n_{i0} . The second example assumes that the mean intensity is large enough so as to invoke the Central Limit Theorem [19], but it considers the improvement that can be achieved when using nonclassical, squeezed light (see, e.g., [30–32]). Under these circumstances, squeezed light produces Gaussian noise statistics with variance $s^2 n_0$, where $s^2 < 1$ is the squeezing factor [33,34]. The number of absorbed photons on a single detector is thus parameterized by the PDFs:

$$f_{N_i}^{\text{sht}}(n_i | n_{i0}) = \frac{(n_{i0} + n_{ib})^{n_i}}{n_i!} \exp[-(n_{i0} + n_{ib})] \quad (19)$$

and

$$f_{N_i}^{\text{sqz}}(n_i | n_{i0}) = \frac{1}{\sqrt{2\pi(s^2 n_{i0} + n_{ib})}} \exp \left[-\frac{(n_i - n_{i0} - n_{ib})^2}{2(s^2 n_{i0} + n_{ib})} \right], \quad (20)$$

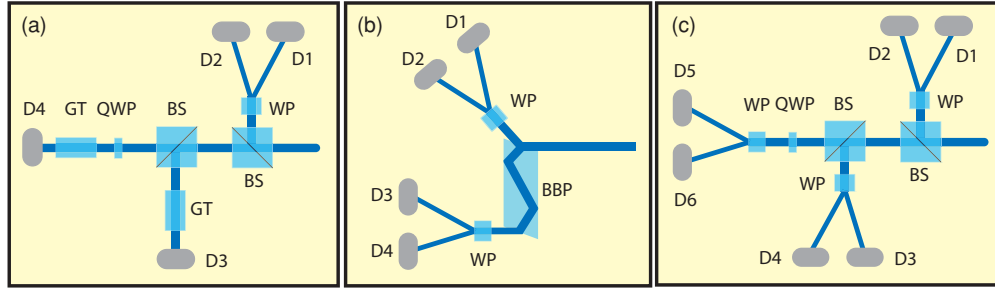


FIG. 2. (Color online) Schematics of three alternative DOAP designs (see text). Notation is as follows: BS is a beam splitter, WP is a Wollaston prism, QWP is a quarter wave plate, GT is a Glan Thompson polarizer, BBP is a broadband prism, and D is a detector.

respectively. $D_i = n_i h \nu_0$ is then the detected intensity on the i th detector.² The additional mean term n_{ib} has been introduced in Eq. (19) [Eq. (20)] to account for other potential additive sources of stray photons, assumed to be independent and Poisson [Gaussian] distributed, such that the joint PDF is also Poisson [Gaussian] distributed. A good discussion of such possible noise sources is given in [35]; however, two simple examples would be a detector dark count or a passive background. Although not necessary, the simplifying assumption that these additional noise sources affect each detector equally such that $n_{ib} = n_b$ is also made. Furthermore, it is reasonable to assume that the noise present on each of the N_D measurements is independent and hence the joint PDF required to calculate the FIM is given by $f_{\mathbf{N}}(\mathbf{n}|\mathbf{n}_0) = \prod_{i=1}^{N_D} f_{N_i}(n_i|n_{i0})$.

Using Eqs. (3), (19), and (20), the FIMs for polarization measurements are given by

$$\mathbb{J}_{\mathbf{w}} = \mathbb{G}^T \mathbb{J}_{\mathbf{D}} \mathbb{G}, \quad (21)$$

where

$$\mathbb{J}_{\mathbf{D}} = \frac{1}{h^2 \nu_0^2} \text{diag} \left[\frac{1}{n_{i0} + n_b} \right], \quad (22)$$

assuming classical shot noise and

$$\mathbb{J}_{\mathbf{D}} = \frac{1}{h^2 \nu_0^2} \text{diag} \left[\frac{1}{s^2 n_{i0} + n_b} \right], \quad (23)$$

for Gaussian noise, whilst

$$\mathbb{G} = \frac{\partial \mathbf{D}_0}{\partial \mathbf{w}} = \mathbb{V} \mathbb{T} \frac{\partial \mathbf{S}}{\partial \mathbf{w}} \quad (24)$$

and $\partial \mathbf{S} / \partial \mathbf{w} = \text{diag}[s_0, s_0, s_0, 1]$. It is immediately apparent from Eqs. (22) and (23) that the use of squeezed light gives an improvement in performance over classical light. Although the potential performance gains from squeezed light have been previously reported in the context of imaging (see, e.g., [33]), this result has not previously been shown for polarimetric studies. It is interesting to note that $\mathbb{J}_{\mathbf{D}} \propto 1/\nu_0^2$. The increase in Fisher information (and associated increase in system accuracy) with lower frequencies arises since this corresponds to more collected photons for a given intensity which, as discussed in [36], corresponds to more independent samples of the stochastic variables. Bayesian FIMs can be

found by performing the Bayesian averaging of Eqs. (22) and (23).

B. Polarimeter architectures

For definiteness, the results presented above are illustrated using three DOAPs existing in the literature. The first of these, as proposed by Azzam [13] and shown schematically in Fig. 2(a), can be easily implemented using only beam splitters, polarizers and waveplates. The detectors in turn project the incident Stokes vector onto horizontal, vertical, linear 45°, and right-circular polarized states, and hence has the instrument matrix

$$\mathbb{T}_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \quad (25)$$

Any noise in the intensity measurements is amplified during data processing to extract, for example, the Stokes parameters, the extent of which is often measured using the condition number of the instrument matrix \mathbb{T} , defined as $\kappa_T = \|\mathbb{T}\|_F \|\mathbb{T}^{-1}\|_F$ where $\|\dots\|_F$ denotes the Frobenius norm. Compain and Drevillon [37] proposed an alternative PSA construction, as shown in Fig. 2(b), in which the prism geometry and the angle of incidence of light onto the first surface are optimized to minimize the condition number of the instrument matrix to a value of 4.48. The associated instrument matrix is

$$\mathbb{T}_2 = \frac{1}{2} \begin{pmatrix} 1 & -0.575 & 0.818 & 0 \\ 1 & -0.575 & -0.818 & 0 \\ 1 & 0.617 & -0.003 & 0.787 \\ 1 & 0.617 & 0.003 & -0.787 \end{pmatrix}. \quad (26)$$

Note that the deviations of \mathbb{V} from the ideal case arising from Fresnel reflection and transmission at the prism entrance surface will be ignored so that the results calculated will be comparable to the alternative DOAPs considered here. The imbalance between reflected and transmitted beams only equates to $\approx 5\%$, however, and discrepancies from reality will thus be small. Instrument matrices with condition numbers smaller than 4.48 are possible [38], but experimental realization of these is complicated since it requires, in general, eight Babinet Soleil compensators. The wavelength dependence of \mathbb{T}_2 is also neglected here.

²Assuming unity quantum efficiency for simplicity.

Finally, a DOAP configuration employing a basis of six distinct measurement states, as given by the instrument matrix

$$\mathbb{T}_3 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix}, \quad (27)$$

was recently proposed by Lara and Paterson [39] and is schematically shown in Fig. 2(c). This DOAP architecture was shown to possess polarization-independent noise characteristics in the Stokes parameters in the presence of a combination of Gaussian thermal noise and signal-dependent Poisson noise.

VI. RESULTS AND DISCUSSION

Using Eqs. (10) and (21)–(27), $\mathbb{J}_{\mathbf{u}}$ can be calculated and hence so too can the accuracy before Bayesian averaging $\mathcal{A}_u = \mathcal{A}_u(\mathbf{u})$. Due to the similarity of the form of the FIM for Gaussian and Poisson noise, restriction is now made to Poisson noise only. Numerical calculations were performed for each DOAP configuration assuming incident light with a wavelength of 405 nm. A total mean photon count of 10^4 and a zero-background count were further assumed. The resulting state-dependent PSA accuracy is shown in Fig. 3. Whilst these plots are formed via direct evaluation of the formulas given in the preceding theory, Monte Carlo simulations (in which the accuracy was calculated from the covariance matrix of simulated random data) are in good agreement.

With reference to Fig. 3 and Eqs. (25)–(27), it is worth mentioning that, for a particular polarimeter architecture, the best accuracy is achieved when measuring totally polarized states that equalize the intensity measured in each polarimeter arm; a result which also holds for general PSA configurations. However, accuracy is seen to decrease with the degree of polarization (i.e., toward the center of the Poincaré sphere), a trend which would be expected and has been discussed in [21].

It is also noted that the PSA architecture proposed by Lara and Paterson is not seen to give a constant accuracy over the surface of the Poincaré sphere, in apparent contradiction

with [39]. This discrepancy, however, arises due to the use of a different metric. The metric proposed in [39] is equivalent to $\text{tr}[\mathbb{J}^{-1}]$, or so-called *A* optimality. However, the metric proposed in this work takes greater account of the cross correlations present in parameter estimates. The six-arm DOAP does, however, still exhibit greater uniformity over the surface of the Poincaré sphere, resulting from a greater sampling of Poincaré space and hence increased redundancy in the experimental data. Sampling of the Poincaré sphere and hence precision can naturally be extended by further increasing the number of detection arms, however so doing has implications on the signal-to-noise ratio in each detector due to the reduced incident intensity (see [21] for further discussion).

If it were known *a priori* that some particular polarization state was more likely to be measured, additional accuracy gains could be made [as described by the \mathbb{J}_w^p term of Eq. (4)] by appropriate system design [21].

Figure 4 shows the variation of accuracy and efficiency of observation (after averaging) as a function of the number of absorbed photons for each DOAP configuration. The accuracy is seen to improve as the number of detected photons increases. Infinite accuracy is thus, in principle, possible in polarimetry if enough photons are detected. A similar conclusion was reached in terms of localization accuracy for two point objects [40]. Additionally, the efficiency falls as photon numbers increase. This essentially arises since there is a redundancy in the information which each photon in a beam carries with regard to their polarization.

Whilst a relatively low number of photons were considered when calculating the data in Fig. 4, it should be observed that both accuracy and efficiency of observation are monotonic functions of the mean photon count. Consequently, trends inferred from Fig. 4 are valid for Poisson noise in which estimator efficiency is only achieved asymptotically [36]. Furthermore, if Gaussian noise described by Eq. (20) was assumed in numerical calculations, plots of identical functional form would follow. Due to the exact efficiency of the maximum-likelihood estimator in Gaussian noise, Fig. 4 is valid even at such low photon counts.

The results given in this article can be practically employed in several ways. For example, since the definition of resolution developed constitutes a fundamental statistical limit, it can be

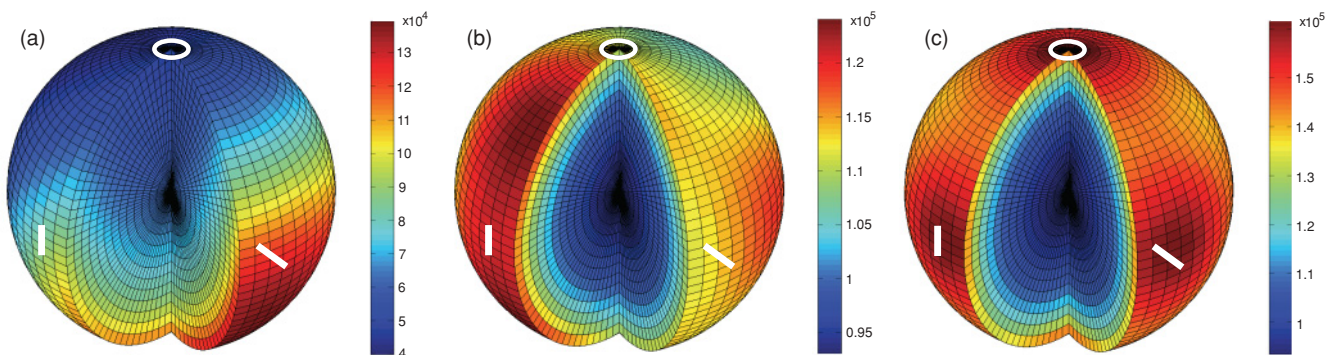


FIG. 3. (Color online) Poincaré diagrams showing the polarization-state dependence of accuracy, before Bayesian averaging, for each DOAP configuration shown in Fig. 2. Simulation parameters used were $n_0 = 10^4$, $n_b = 0$, $\lambda_0 = c/\nu_0 = 405$ nm, and $c = 0.9$. White markers denote the state of polarization at different points on the Poincaré sphere and are shown for reference.

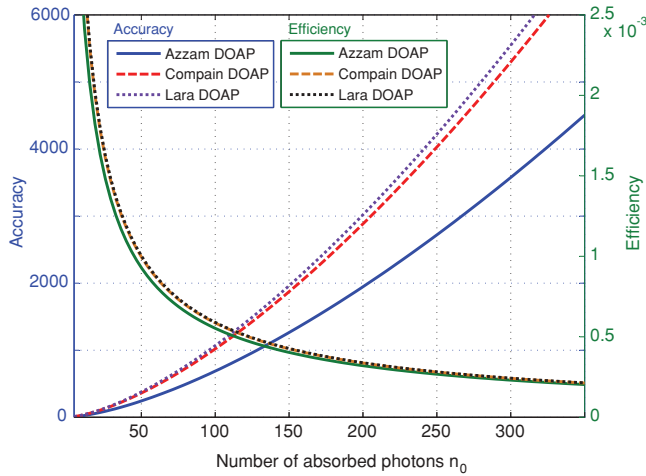


FIG. 4. (Color online) Variation of accuracy and efficiency of observation with mean photon count for different DOAP configurations (see text). Simulation parameters are the same as for Fig. 2

used as a benchmark for newly developed polarimetric systems or signal-processing algorithms. One must thus distinguish between the theoretical performance limit and that practically achievable using a given system [cf. Eq. (2)]. Both can be evaluated using the performance metrics given above [e.g., Eqs. (8), (12), and (18)] but, for the latter case, the FIM should be replaced with the achievable covariance matrix \mathbb{K} . Consider, for instance, the experimental evaluation of the polarization resolution in Stokes space achievable with a novel DOAP. For any given incident polarization state \mathbf{S} , an ensemble of data sets must be taken; that is, multiple measurements under fixed experimental conditions. Inferring the parameter vector $\hat{\mathbf{w}}_i$ from the i th set of data, the covariance matrix can be estimated using the standard statistical formulas:

$$\boldsymbol{\mu}(\mathbf{w}) = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{w}}_i, \quad (28)$$

where M is the number of measurements taken and

$$\hat{\mathbb{K}}_{\mathbf{w}}(\mathbf{S}) = \frac{1}{M-1} \sum_{i=1}^M (\hat{\mathbf{w}}_i - \boldsymbol{\mu})(\hat{\mathbf{w}}_i - \boldsymbol{\mu})^T. \quad (29)$$

Naturally, if the noise present in the system were known to be independent of the incident polarization state, only a single state of polarization would need to be used. However, given the possibility of signal-dependent noise, a full system characterization requires the use of a polarization-state generator, so that the sample mean and covariance, given by Eqs. (28) and (29), can be determined for all input polarization states \mathbf{S} . Given the experimental covariance as a function of position in Stokes space, it is a simple matter to substitute it numerically into the desired metric.

Benchmarking, however, in which the experimental performance is compared to the theoretical limit, requires knowledge of the full noise statistics as encapsulated in the PDF $f(\mathbf{D}|\mathbf{w})$. However, since the mapping from intensity vector \mathbf{D} to the parameters of interest (\mathbf{w} in this example) is known, it is sufficient to consider the PDF $f(\mathbf{D}|\mathbf{D}_0)$ [cf. Eq. (24)]. A theoretical noise model of the system means the PDF can be directly substituted into the metric definitions as above; however, in the absence of such a model it is necessary to determine the statistics experimentally. The experimental configuration will determine whether it is easier to estimate $f(\mathbf{D}|\mathbf{w})$ or $f(\mathbf{D}|\mathbf{D}_0)$. Regardless, one method to measure the desired PDF³ is to histogram the measured intensities for a fixed input power (or polarization state), as found from repeated experiments. Since the derivatives of the likelihood function are needed to calculate the FIM [i.e., $\partial \ln f(\mathbf{D}|\mathbf{D}_0)/\partial \mathbf{D}_0$ or, alternatively, $\partial \ln f(\mathbf{D}|\mathbf{w})/\partial \mathbf{w}$], it is again necessary to take repeated measurements for different illumination conditions whereby the derivatives can be approximated using discrete differences. Given the capability to benchmark systems, the presented metrics immediately present a basis for optimization and system design.

Finally, it is worth noting that, although the formulas have been presented in the context of estimating a position in Stokes or Poincaré space, the formalism can be simply extended to calculate the accuracy of inferred polarization properties such as the degree of polarization or dichroism of a sample, allowing for more rigorous noise analysis in polarimetric systems.

³Note that no claim as to the optimality of this approach is being made here, but it is instead presented as a simple method to illustrate the principle.

- [1] A. C. S. Readhead *et al.*, *Science* **306**, 836 (2004).
 [2] R. Oldenbourg, *Nature (London)* **381**, 811 (1996).
 [3] A. Bnire, F. Goudail, D. Dolfi, and M. Alouini, *J. Opt. Soc. Am. A* **26**, 1678 (2009).
 [4] B. Sick, B. Hecht, and L. Novotny, *Phys. Rev. Lett.* **85**, 4482 (2000).
 [5] J. Ellis and A. Dogariu, *Phys. Rev. Lett.* **95**, 203905 (2005).
 [6] A. Shields, *Science* **297**, 1821 (2002).
 [7] J. Desroches, D. Pagnoux, F. Louradour, and A. Barthélémy, *Opt. Lett.* **34**, 3409 (2009).
 [8] U. Seljak and M. Zaldarriaga, *Phys. Rev. Lett.* **78**, 2054 (1997).
 [9] M. R. Andrews, P. P. Mitra, and R. de Carvalho, *Nature (London)* **409**, 316 (2001).
 [10] D. G. M. Anderson and R. Barakat, *J. Opt. Soc. Am. A* **11**, 2305 (1994).
 [11] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light* (Elsevier, Amsterdam, 1987).
 [12] J. S. Tyo, *Appl. Opt.* **41**, 619 (2002).
 [13] R. M. A. Azzam and N. M. Bashara, *J. Mod. Opt.* **29**, 685 (1982).
 [14] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946).
 [15] C. Rao, *Bull. Calcutta Math. Soc.* **37**, 81 (1945).
 [16] R. A. Fisher, *Philos. Trans. R. Soc. London* **222**, 309 (1922).
 [17] R. A. Fisher, *Proc. Cambridge Philos. Soc.* **22**, 700 (1925).
 [18] M. Brookes, The matrix reference manual. [<http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html>], 2005.

- [19] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* (Addison-Wesley, Reading, 1991).
- [20] H. L. van Trees, *Detection, Estimation, and Modulation Theory: Part I* (John Wiley & Sons, New York, 1968).
- [21] M. R. Foreman, C. Macías Romero, and P. Török, *Opt. Express* **16**, 15212 (2008).
- [22] G. E. Keiser, *Opt. Fiber Technol.* **5**, 3 (1999).
- [23] International Organization for Standardization and European Computer Manufacturers Association. Optical interfaces for multichannel systems with optical amplifiers. ITU-T Recommendation G. 692, International Telecommunication Union, 1998.
- [24] I. J. Cox and C. J. R. Sheppard, *J. Opt. Soc. Am. A* **3**, 1152 (1986).
- [25] M. R. Foreman, S. S. Sherif, and P. Török, *Opt. Express* **15**, 13597 (2007).
- [26] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948); **27**, 623 (1948).
- [27] L. Szilard, *Z. Phys.* **53**, 593 (1929).
- [28] L. Brillouin, *J. Appl. Phys.* **22**, 334 (1950).
- [29] L. Brillouin, *J. Appl. Phys.* **22**, 338 (1950).
- [30] N. Korolkova, G. Leuchs, R. Loudon, T. C. Ralph, and C. Silberhorn, *Phys. Rev. A* **65**, 052306 (2002).
- [31] W. P. Bowen, N. Treps, R. Schnabel, and P. K. Lam, *Phys. Rev. Lett.* **89**, 253601 (2002).
- [32] J. Milanovic, M. Lassen, U. L. Andersen, and G. Leuchs, *Opt. Express* **18**, 1521 (2010).
- [33] B. E. A. Saleh and M. C. Teich, *Phys. Rev. Lett.* **58**, 2656 (1987).
- [34] V. Delaubert, N. Treps, C. Fabre, A. Maître, H. A. Bachor, and P. Réfrégier, *Europhys. Lett.* **81**, 44001 (2008).
- [35] A. Bénéière, F. Goudail, M. Alouini, and D. Dolfi, *J. Europ. Opt. Soc. Rap. Public.* **3**, 08002 (2009).
- [36] H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, *J. Opt. Soc. Am. A* **12**, 834 (1995).
- [37] E. Compain and B. Drevillon, *Appl. Opt.* **37**, 5938 (1998).
- [38] S. N. Savenkov, *Opt. Eng.* **41**, 965 (2002).
- [39] D. Lara and C. Paterson, *Opt. Express* **17**, 21240 (2009).
- [40] S. Ram, E. S. Ward, and R. J. Ober, *Proc. Natl. Acad. Sci. USA.* **103**, 4457 (2006).