Stokes polarimeter optimization in the presence of shot and Gaussian noise

David Lara and Carl Paterson
The Blackett Laboratory, Imperial College London, SW7 2BW, UK
*d.lara@imperial.ac.uk

Abstract: An error minimization method is presented for Stokes polarimeters applicable when the detected signals are affected by a combination of shot and Gaussian noise. The expectation of the Stokes vector variance is used as a performance measure. This measure is compared with the condition number of a polarization state analyzer matrix that is commonly used as a figure of merit. We show that a polarimeter with the minimum condition number is not necessarily optimal. The approach is used to optimize existing prism based polarimeters giving improvements in the performance when shot-noise cannot be neglected.

© 2009 Optical Society of America

OCIS codes: (260.5430) Polarization; (030.4280) Noise in imaging systems; (120.2130) Ellipsometry and polarimetry.

References and links
1. J. S. Tyo, “Design of optimal polarimeters: maximization of signal-to-noise ratio and minimization of systematic error,” Appl. Opt. 41(6), 619–630 (2002).
2. A. D. Martino, E. Garcia-Cauriel, B. Laude, and B. Drevillon, “General methods for optimized design and calibration of Mueller polarimeters,” Thin Solid Films 455–456, 112–119 (2004).
3. D. S. Sabatke, M. R. Descour, E. L. Dereniak, W. C. Sweatt, S. A. Kemme, and G. S. Phipps, “Optimization of retardance for a complete Stokes polarimeter,” Opt. Lett 25(11), 802–804 (2000).
4. K. M. Twietmeyer, R. A. Chipman, A. E. Elsner, Y. Zhao, and D. VanNasdale, “Mueller matrix retinal imager with optimized polarization conditions,” Opt. Express 16(26), 21,339–21,354 (2008).
5. A. E. Elsner, S. A. Burns, J. J. Weiter, and F. C. Delori, “Infrared imaging of sub-retinal structures in the human ocular fundus,” Vision Research 36(1), 191–205 (1996).
6. F. C. Delori and K. P. Pflibsen, “Spectral reflectance of the human ocular fundus,” Appl. Opt. 28(6), 1061–1077 (1989).
7. M. R. Foreman, C. M. Romero, and P. Török, “A priori information and optimisation in polarimetry,” Opt. Express 16(19), 15,212–15,227 (2008).
8. R. M. A. Azzam, “Division-of-amplitude photopolarimeter (DOAP) for the simultaneous measurement of all Stokes parameters of light,” Opt. Acta 29(5), 685–689 (1982).
9. E. Compain and B. Drevillon, “Broadband division-of-amplitude polarimeter based on uncoated prisms,” Appl. Opt. 37(25), 5938–5944 (1998).
10. E. Compain, S. Poirier, and B. Drevillon, “General and self-consistent method for the calibration of polarization modulators, polarimeters and Mueller-matrix ellipsometers,” Appl. Opt. 38(16), 3490–3502 (1999).
11. B. Laude-Boulesteix, A. de Martino, B. Drevillon, and L. Schwartz, “Mueller polarimetric imaging system with liquid crystals,” Appl. Opt. 43(14), 2824–2832 (2004).
12. D. Lara and C. Dainty, “Axially resolved complete Mueller matrix confocal microscopy,” Appl. Opt. 45(9), 1917–1930 (2006).
13. PerkinElmer, “Avalanche Photodiodes: A User’s Guide,” Technical information, PerkinElmer Optoelectronics (2006).
14. Hamamatsu, “Characteristics and use of Si APD (Avalanche Photodiode),” Technical Information SD-28, Hamamatsu Photonics K. K., Solid Division (2004).
15. R. J. McIntyre, “Multiplication noise in uniform avalanche diodes,” IEEE Trans. Electron Dev. 13(1), 164–168 (1966).
1. Introduction

Stokes vector polarimeters and polarization sensitive imaging devices can record important information from a beam of light that is not accessible to conventional intensity or phase contrast detection devices. Most of the existing optimization methods for Stokes polarimeters have focused on independent Gaussian noise [1–3], such as is appropriate for dominant detector readout noise. However, advances in detection technology and polarimeter design, theory, and calibration are spreading the implementation of polarimetry techniques to applications where low light levels and high acquisition rates often impose a limit on the performance of the device through signal dependent (shot) noise. For example, in a polarization sensitive ophthalmoscope [4], the amount of light that can be used is limited for the safety of the subject, and the return is typically a small fraction of the light used for illumination, of the order of $10^{-6}$. Furthermore, the acquisition rates have to be sufficiently fast to avoid eye-movement artifacts. The light-budget available to produce a polarization image with a scanning laser ophthalmoscope is typically in the order of 10 nW [5, 6], after the signal has been analyzed with a division-of-amplitude polarimeter (DOAP), and pixel acquisition sampling rates are of several MHz. If light coming back from the sample is not split with a DOAP, the data sampling rates need to increase to enable adequate pixel registration of the different polarization signals.

The overall optimization of a system to measure the effect of a sample on polarization (e.g. Mueller matrix polarimeter and data processing) should take into account the polarization state generator (PSG) as well as the polarization state analyzer (PSA), and any a-priori knowledge of the effect of the sample [7]. However, when there is little quantitative data available about the polarization features of the object of study, or when there is no access to control the illumination system, it is appropriate to consider PSAs that are capable of measuring any state of polarization.

In this article we evaluate three PSA schemes in the presence of a combination of shot noise and thermal noise using a polarimeter optimization cost function based on the Stokes parameters. Our focus is on division of amplitude polarimeters (DOAP) [8] that can simultaneously measure the four or more intensity values needed to estimate the Stokes vector of a beam of light, but the approach is valid for any kind of Stokes vector polarimeter. We then use the approach to optimize a DOAP based on a prism design first proposed by Compain and Drevillon [9].

In this paper we use a realistic model of a combination of an avalanche photodiode (APD) and a transimpedance amplifier in the optimization of Stokes polarimeters where the sources of noise are neither pure Gaussian nor pure Poissonian, but a combination of the two. The approach, however, is not limited to the choice of noise model we use here, and could be used with different noise characteristics appropriate for other detectors like photomultiplier tubes (PMT) and silicon PIN detectors, which may be more suitable for different light budgets or polarimeter applications. Systematic errors due to misalignment of elements or imperfect components can be eliminated with existing robust calibration methods, that are available in single pass [10, 11] and double pass configurations [12], and for this reason are not considered here.
2. Background

Stringent illumination conditions make the avalanche photodiode (APD) a suitable detector for a laser scanning ophthalmoscope. APDs have three main sources of noise [13, 14]: 1) fluctuations in the dark surface current, 2) fluctuations in the dark leakage current which are amplified by the gain of the APD, and 3) excess (shot) noise due to the statistical nature of the avalanche multiplication process. This last type of noise depends on the optical signal strength and the gain of the APD, and can be much higher than the other two, which results in shot noise limited operation of the APD. After the avalanche multiplication the current produced by the APD is typically fed to a transimpedance amplifier and therefore is affected by the amplifier’s thermal (Gaussian) noise. The exact mechanism of the production of noise in a APD-amplifier module falls beyond the scope of this paper, but it is well understood and documented [13–15]. We may note, however, that the combination of all these sources of noise results in a strong dependence of the overall signal to noise ratio (SNR) on the gain of the APD. This SNR reaches an optimum value approximately when the shot noise equals the thermal noise [13], hence neither should be neglected. Figure 1 shows the SNR as a function of the APD gain in a realistic detector suitable for a scanning laser ophthalmoscope.

![SNR vs APD Gain](image)

Fig. 1. Signal to noise ratio of an APD and transimpedance amplifier combination operating at 8MHz sampling rate and with 5 nW of optical power falling on the detector. Realistic model of APD Perkin Elmer C30902E and a transimpedance amplifier with noise current of 3pA/Hz$^{1/2}$.

3. Method

3.1. Cost function

For a given Stokes vector $\mathbf{S}$, with components $s_j$ ($j = 0, 1, 2,$ and $3$), the intensity $i_k$ falling on each detector of a polarimeter, in the absence of noise, is given by

$$i_k = D_{kj} s_j,$$

(1)

where $D_{kj}$ is an element of the the PSA instrument matrix $\mathbf{D}$. This matrix is of size $m \times 4$, where $m$ is the number of detected signals used to measure a Stokes vector. From a set of intensity measurements $i_k$, we estimate the Stokes vector using a least squares reconstructor by means of $\mathbf{F}$, the Moore-Penrose pseudoinverse [16] of $\mathbf{D}$. So, in the absence of noise

$$\hat{s}_j = F_{jk} i_k,$$

(2)

where $F_{jk}$ is an element of $\mathbf{F}$.

Assuming that noise in each detected signal is statistically independent from the others, the noise variance on the estimated Stokes parameters is given by

$$n^2_{s_j} = F_{jk}^2 n^2_k.$$ 

(3)
The noise variance \( n^2 \) on each detected signal is a function of the signal strength \( i_k \) when shot-noise in the system cannot be neglected. For this reason, in general, the noise variance \( n^2 \) on each Stokes parameter depends on the state of polarization being measured.

We define the square error in the estimation of any given Stokes vector as the sum of the square errors of its parameters, and we denote it by

\[
\varepsilon^2_S = \sum_{j=0}^{j=3} (s_j - \hat{s}_j)^2 + (n_s)^2; \tag{4}
\]

where \( \hat{s}_j \) is taken from Eq. (2). The term \( (s_j - \hat{s}_j)^2 \) represents the reconstruction square error in the absence of noise, and is only non-zero in cases when \( D \) is less than full rank.

The expected variance over an ensemble of Stokes vectors is given by

\[
\sigma_{PSA}^2 = \int \varepsilon^2_S p(S) d^nS, \tag{5}
\]

where \( p(S) \) is the probability distribution for the Stokes vectors in the ensemble. The square root of the expected variance is then taken to define a cost function for optimization of the PSA.

The probability distribution for the Stokes vectors depends on a-priori knowledge that may be available about the beam of light to analyze, the Mueller matrix to measure, or the PSG implemented. For example, if we knew that all light falling on our PSA was completely and linearly polarized, we would choose a distribution of states of polarization that lie on the equator of the Poincaré sphere to evaluate the PSA through the cost function in Eq. (5). In the examples below we choose a uniform distribution over the surface of the Poincaré sphere as it is equivalent to having minimum prior knowledge of the state of polarization to be measured.

It is possible to use probability distributions that include points inside the Poincaré sphere (i.e. not fully polarized) in \( p(S) \), but such should only be used with care. In beam like fields, partial polarization or unpolarized light arise when the measured Stokes vector is an average of different states of completely polarized light. This polarization averaging may well be spectral (wavelength), spatial or temporal, but will inevitably depend on the instrument itself, as much as on the nature of the object/beam being measured. In a scanning microscope, for example, measured depolarization may arise if each set of detected signals corresponds to the spatial average of different positions on the sample along a scanning direction of the instrument. If the scanning is made more rapidly and the integration time kept the same, depolarization will increase in a non-homogenous sample, even though the sample had not changed. Under coherent imaging conditions [17], measured depolarization is an artifact of the instrument, which can be used as a system check to determine if all sampling rates (temporal, spatial, spectral, etc.) are adequate.

Here we consider only polarized light in the cost function, although this does not prevent the estimation of partially polarized states with the linear reconstructor. We note that mean error variance is not the only possible metric. For example Goudail [18] uses maximum variance, relating this directly to condition number under conditions of pure Poisson noise. The choice of cost function will depend on the intended use of the polarimeter, and in an imaging application, minimizing the expected variance is equivalent to minimizing the noise power in the image.

### 3.2. Numerical implementation

In this paper the cost function is evaluated numerically. The particular method we choose is to sample the states of polarization along a spiral locus around the Poincaré sphere with the spiral locus and azimuthal sampling chosen to give approximately uniform distribution of samples over the sphere (Fig. 2). By mapping samples to a one-parameter locus this method also makes
it straightforward to show the behavior over the whole sphere on a simple plot, as will be seen later in Figs. 3–5.

![Poincaré sphere](image)

Fig. 2. Sampling of states of polarization represented on the surface of the Poincaré sphere. Uniformly distributed samples are taken along a right-handed spiral locus that starts at the south pole and finishes at the north pole, with 128 samples per revolution (For clarity the figure shows only 24 revolutions, though 128 were used in the numerical evaluation). Parameters $u$ and $\theta$ are uniformly distributed within the intervals shown.

### 4. Results

We use the approach to evaluate the performance of three known PSA configurations in a retinal imaging type application. For the comparison, we assume 30 nW of optical power entering the polarimeter, detectors with the characteristics of a commercial avalanche photodiode (Perkin Elmer C30902E) coupled with a transimpedance amplifier with realistic noise current of 3pA/Hz$^{1/2}$, and measurement sampling at 8MHz. We keep the APD gain fixed at a value of 100 as it gives the maximum SNR when 5 nW are used in the detector model (see Fig. 1), and is the expected light level falling on each detector. The noise on each reconstructed Stokes parameter, $n_{S_j}$ (square root of noise variance), are plotted in Figs. 3(b)–5(b), and the corresponding values of $\sigma_{PSA}$ are given in Table 1. The noise plots are shown for all points along the spiral locus of Fig. 2, moving from the south pole to the north pole. For clarity, the figures show only 24 revolutions, though 128 were used in the numerical evaluation that appears in Table 1.

#### 4.1. Objective comparison of different PSAs

We first consider a PSA comprising beam-splitters, linear polarizers and waveplates [12, 19] (henceforth referred to as BS polarimeter). Figure 3 shows the performance for a PSA similar to that described in [12] but with the addition of an ideal quarter waveplate before the 45° linear polarizer and with the transmittance–reflectance ratio of the first non-polarizing beam-splitter adjusted to 2/3:1/3 to distribute unpolarized light equally among the four detectors, thereby improving its performance and establishing a better comparison. The instrument matrix [Eq.
for this PSA is

\[
\mathbf{D}_{\text{BS}} = \begin{bmatrix}
0.1978 & 0.1978 & 0 & 0 \\
0.1978 & -0.1978 & 0 & 0 \\
0.1514 & 0.0102 & 0.1497 & 0.0206 \\
0.1508 & -0.0014 & 0.0205 & 0.1494
\end{bmatrix}.
\]

(6)

The last three elements on each row of the instrument matrix are coordinates of eigenvectors of the polarization detectors in the PSA. These vectors, normalized to unit magnitude, are plotted as bold dots on the surface of the Poincaré spheres of Figs. 3(a), 4(a), and 5(a).

Note in Fig. 3(b) that the noise on a given Stokes parameter is largest when that parameter dominates in the Stokes representation of the state of polarization and the Stokes vector is closer to an eigenstate of the PSA. For example, \(n_{S_3}\) in Fig. 3 is larger where the Stokes vectors are closer to the north pole (where the value of \(S_3\) is dominant) and near to a polarization detector in the PSA. This is a consequence of having a significant contribution by shot noise, which means that, inevitably, as the signal becomes larger so does the noise. This is critical when trying to detect small fluctuations in the polarization signal, and appropriate optimization is paramount.

Figure 4 shows the results for the prism described by Compain and Drevillon [9] (henceforth Compain polarimeter). The shape of the Compain prism had been designed to minimize the condition number of the instrument matrix. Later, we show that the noise cost function can be made smaller by choosing a different geometry of the prism. We use here the theoretical PSA matrix in Eq. (17) of Ref. [9], without the artificial gain factor of 1000 that Compain and Drevillon used in their publication in order to compare with their experimental matrices.

\[
\mathbf{D}_{\text{Compain}} = \begin{bmatrix}
0.1875 & -0.1079 & 0.1534 & 0 \\
0.1875 & -0.1079 & -0.1534 & 0 \\
0.1950 & 0.1203 & 0 & 0.1535 \\
0.1950 & 0.1203 & 0 & -0.1535
\end{bmatrix}.
\]

(7)

Finally, Fig. 5 shows results for an ideal six-detector polarimeter. The six-detector polarimeter was heuristically modeled as to distribute the optical power of a globally unpolarized beam [20] equally among the six detectors. The instrument matrix for this polarimeter is given
Fig. 4. (Color online) Compain polarimeter. (a) and (b) as per Figure 3.

by

\[
D_{\text{det}} = \frac{1}{6} \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & -1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & -1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & -1
\end{bmatrix}.
\] (8)

(The arrangement is easy to implement with simple ideal components: the light can be split into three equal components using non-polarizing beam splitters one of which is then projected into horizontal and vertical linear components, another into +45° and −45° linear components and the third into left- and right-circular polarization states with the aid of Wollaston prisms and a quarter-wave plate.)

We note that in the six-detector polarimeter (Fig. 5) the noise on the Stokes parameters does not depend on the state of polarization being measured and that this PSA gives the best cost performance of the three systems compared.

Fig. 5. (Color online) Six-detector polarimeter. (a) and (b) as per Fig. 3.
4.2. Optimization of a PSA design

The optimization method we present is used to improve the performance of the original Com- pain prism. We allowed the PSA prism’s shape to vary without constraining it to be a parallelepiped. The parameters that defined the model of the prism were the material (refractive index and absorption), the three shape angles $\phi_1$, $\phi_2$, and $\phi_3$ illustrated in Fig. 6, and the angle of incidence of the beam ($\phi_0$) on the prism. As in the Compain PSA, an ideal polarizing beam-splitter was placed at each exit port A and B, shown in Fig. 6. The orientations of these last beam-splitters were initially left as free parameters, but the optimization always returned 45$^\circ$ for both beam-splitters. The path length inside the prism was fixed at 10 cm to allow fair comparison between materials of different absorption. We performed optimizations for three different types of glass that resulted in three different prism geometries. The cost functions of the optimized prisms are shown in Table 1, together with those for the three PSAs compared in Section 4.1. The first optimized prism (Optim1) uses the same glass parameters as the Com- pain prism: Corning glass E00-046 (refractive index (RI) = 1.812, absorption constant $\alpha$=1.123 m$^{-1}$). The noise cost function was reduced despite the new prism having a slightly larger condition number than the Compain prism (see Table 1). The second prism (Optim2) uses a different glass, Shott SF57HHT (RI = 1.872, $\alpha = 0.548$ m$^{-1}$), which resulted in a even smaller cost function. Finally, we included the refractive index of the material as a free parameter, with an ideal zero absorption constant (Optim3). Note that reducing the overall pathlength inside the prism reduces the impact of absorption, although the extent to which this is possible depends on the required optical geometry. By including these material effects in the optimization, the method presented can be used to evaluate the suitability of different materials for a PSA. For the sake of comparison, Figure 7 shows the noise on the Stokes parameters calculated for the Optim3 prism. The resulting parameters for the optimized prism are: RI=2.11, $\phi_0 = 81.19^\circ$, $\phi_1 = 78.53^\circ$, $\phi_2 = 52.85^\circ$, and $\phi_3 = 53.54^\circ$. The instrument matrix for this optimized polarimeter is

$$D_{\text{Optim3}} = \begin{bmatrix} 0.2337 & -0.1259 & 0.1969 & 0 \\ 0.2337 & -0.1259 & -0.1969 & 0 \\ 0.2419 & 0.1427 & 0 & -0.1953 \\ 0.2419 & 0.1427 & 0 & 0.1953 \end{bmatrix}.$$  

(9)

Since this is the matrix of a DOAP the sum the elements on the first column corresponds to the efficiency of light collection of the polarimeter, in this case 95%. This optimized design has the lowest expected error of the polarimeters considered (Table 1), including the ideal six-detector polarimeter, which has 100% light efficiency and a smaller condition number.

Fig. 6. Shape parameters for the optimized polarimeter.
Fig. 7. (Color online) Noise on the Stokes parameters for optimized PSA (Optim3) with 30 nW optical power at 8MHz.

Table 1. Cost function and condition number ($\delta$) of the polarimeters evaluated for 30 nW and 10 nW input optical power sampled at 8MHz.

| PSA                 | $\sigma_{\text{PSA}}(30\text{ nW})$ | $\sigma_{\text{PSA}}(10\text{ nW})$ | $\delta$  |
|---------------------|-------------------------------------|-------------------------------------|-----------|
| Optim3 (RI free par.) | 2.77 nW                             | 1.80 nW                             | 1.03 $\sqrt{3}$ |
| Six Detector        | 2.79 nW                             | 1.82 nW                             | $\sqrt{3}$ |
| Optim2 (SF57HHT)    | 2.85 nW                             | 1.85 nW                             | 1.13 $\sqrt{3}$ |
| Optim1 (E00-046)    | 2.92 nW                             | 1.90 nW                             | 1.16 $\sqrt{3}$ |
| Compain [9]         | 3.15 nW                             | 2.05 nW                             | 1.02 $\sqrt{3}$ |
| BS                  | 4.27 nW                             | 2.79 nW                             | 1.65 $\sqrt{3}$ |

5. Conclusion

In the presence of shot-noise, minimizing the condition number does not necessarily give the best performance in terms of minimizing the measurement error on the Stokes vector. When shot noise and Gaussian noise are non-negligible it is appropriate to optimize using the expectation of the Stokes vector variance as a cost function. We used this basis to compare the performances of three PSAs — a simple 4-detector beamsplitter PSA, the Compain prism PSA and an ideal six-detector PSA — taking account of both Gaussian and shot noise at levels appropriate for a retinal imaging application. Of these, the six-detector polarimeter with ideal 100% light collection showed the best performance. In further optimization of the prism type PSA to minimize the expected Stokes vector variance, we showed that performance better than the ideal six-detector polarimeter is achievable despite some loss of optical efficiency and a slightly higher condition number.

Goudail previously showed that optimizing the condition number of the instrument matrix is equivalent to minimizing the maximum variance of a Stokes polarimeter under conditions of pure Poisson noise statistics. However, when the noise comprises both Gaussian and shot components, typical of many real applications, this approach may not be ideal. Moreover, in an imaging application, expected variance may be a more suitable metric than maximum variance, since it relates directly to the noise power in the image. The approach we present is quite general and although we considered only shot and Gaussian noises, the approach is also applicable in the presence of other types of random measurement noise as it is based on the error on the
Stokes parameters rather than a specific noise model. Similarly, although we assumed minimal prior knowledge of the Stokes vector to be measured, the method could be applied incorporating prior knowledge by modifying the expectation distribution in the cost function. Such an approach may have particular benefit in imaging of weakly polarizing objects.

Acknowledgements

This work was carried out with financial support of the Wellcome Trust. C. Paterson acknowledges support from the Royal Society.