Optimizing Mueller polarimetry in noisy systems through over-determination

applied optics

Optimizing Mueller polarimetry in noisy systems through over-determination

H. Philpott,[1,4] E. Garcia-Caurel,[2] O. Guaitella,[3] and A. Sobota[1,5]

[1]Eindhoven University of Technology, Eindhoven, The Netherlands
[2]LPICM, CNRS, École Polytechnique, Institut Polytechnique de Paris, 91120, Palaiseau, France
[3]LPP, École Polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau, France
[4]e-mail: h.w.philpott@tue.nl
[5]e-mail: a.sobota@tue.nl

Received 8 July 2021; revised 9 September 2021; accepted 9 September 2021; posted 13 September 2021 (Doc. ID 435085); published 21 October 2021

Mueller polarimetry measurements are increasingly being used to image highly dynamic and short-lived phenomena such as plasma discharges. For phenomena such as these, exposure times below 1 μs must be used. Unfortunately, these low exposure times significantly reduce the signal-to-noise ratio, making accurate and consistent measurements difficult. To overcome this limitation, we investigated increasing the number of Stokes vectors produced from a polarization state analyzer and polarization state generator, a process known as over-determination. To conduct our analysis, we used results from physical experiments using Stokes vectors generated by liquid crystal variable retarders. These results were then verified using data from simulations. First, we conclude that increasing the degree of over-determination is a simple and effective way of dealing with this noise; however, we also convey that choosing the best scheme is not an entirely trivial process. Second, we demonstrate that over-determination gives rise to hitherto inaccessible information that allows for the quantification of statistical noise and, crucially, the pinpointing of the origin of systematic error, a highly beneficial process that has been lacking until now.

https://doi.org/10.1364/AO.435085

1. INTRODUCTION

Mueller polarimetry is an experimental technique used to determine the Mueller matrix of a sample [1,2]. This is achieved by sequentially illuminating the sample with polarized light emerging from a polarization state generator (PSG) and then analyzing the light reflected or transmitted through the sample with a polarization state analyzer (PSA) [3,4]. When used in concert, these two components make it possible to deduce the full description of how the sample alters the polarization properties of light; this information is then stored in a 4×4 matrix known as a Mueller matrix [5]. Alongside Mueller polarimetry, there is Stokes polarimetry [6], which is often referred to simply as polarimetry. Stokes polarimetry is limited to the calculation of Stokes vectors of light rather than Mueller matrices of samples. One of the advantages of polarimetry is that measurements can be made remotely and quickly, and as a consequence of this, it has found applications in several wide ranging areas such as astronomy and geoscience [7–9]. Mueller polarimetry on the other hand is predominantly employed on static targets, and as such, it has found uses in biomedicine and crystallography [10–13], where the full Mueller matrix can reveal valuable insights about the samples. Nevertheless, Mueller polarimetry can also be used in dynamic settings. This is due to a set of materials whose Mueller matrix is dependent on external factors such as temperature and electric fields, known as photo-elastic and electro-optic crystals, respectively [14,15]. When these modulators are placed in a Mueller polarimeter, the changes in the Mueller matrix can be determined and linked back to the active effect. A prime example of this is shown in a plasma diagnostic technique that exploits the Pockels effect to calculate the electric field of plasma impinging upon an electro-optic crystal [16,17]. Dynamic imaging such as this has a crucial drawback though. As the exposure time is decreased, the signal-to-noise ratio also decreases, raising the difficulty of producing accurate results in short-lived phenomena such as plasma. However, noise plagues measurements regardless of exposure time, so there already exists a large area of research dedicated to reducing the effects of noise in both Stokes and Mueller polarimetry [18–23]. Much of this research draws on the rich mathematical foundation that underpins the Stokes framework, so much so that one can easily become lost in the abstract world of group theory and Minkowski metrics [24–26] to the point where it is easy to forget that Mueller polarimetry is principally an experimental technique, not a mathematical architecture. However, this preponderance of theoretical research has certainly proved useful; in particular, the concept of well-determined and

over-determined measurement schemes forms the starting point of the investigation outlined in this paper. Well-determined and over-determined denote the same concept in both Mueller and Stokes polarimetry, but the specifics are different. Within Stokes polarimetry, the object of calculation is a four-dimensional vector, and therefore four measurements are required to produce a well-determined result; but if more measurements are used, the result is deemed as over-determined. To contrast this with Mueller polarimetry, where there are 16 distinct parameters, a well-determined result requires 16 measurements, and an over-determined result needs more than 16. Currently the bulk of research exploring over-determination is applied to Stokes polarimetry [19,21]. In this paper, we set out to rectify this paucity of information and investigate over-determination in the context of Mueller polarimetry. Our work is split into two areas: first we show the positive effect of over-determination on the stability and accuracy of results, then we explore how the metadata arising through over-determination can allow for the quantification of noise within the measurements. The ramifications of our findings lead us to conclude that over-determining a Mueller matrix to any degree is a worthwhile endeavor even in systems relatively devoid of noise; however, there is still more research needed into the degree of over-determination required and also the relationship between noise and metadata terms needs to be studied further.

2. EXPERIMENTAL SETUP

The experiments were performed on the setup shown in Fig. 1. As the figure shows, the light source was an LED producing incoherent collimated monochromatic red light (625 nm), which passes through the PSG, which comprises a linear polarizer (LP) mounted at 180°, followed by a pair Meadowlark liquid crystal variable retarders (LCVRs) [3] that were mounted at 45° and 0°. Each pair of LCVRs was managed by a Meadowlark control unit that applies a voltage between 0 V and 10 V and a minimum switching time of roughly 500 μs to the liquid crystal, thus altering the retardance of the liquid crystal over a range of 15° to 350°. Once the light passes through the PSG, it travels through the sample and into a collection of two lenses before it passes through the PSA. In an ideal system, these two lenses would not be necessary, as the light would be perfectly collimated throughout the whole setup. However, the light source used was capable of producing collimated light only over distances shorter than this setup, and therefore this sequence of lenses was added to ensure that the light entering the PSA was collimated and that the sample was in focus at the camera. The PSA is made up of the same components as the PSG, except now the order of the components (from left to right) is as follows: two LCVRs mounted at 0° and 45°, then a LP mounted at 180°. Immediately in front of the Andor Istar ICCD camera, there is a lens that magnifies the image before entering the camera's array of detectors, which measures the intensity of the light, producing a 1024 × 1024 pixel image. The camera used for these measurements had possible gate widths ranging from seconds to picoseconds, and also had an adjustable gain ranging from 1 to 4095 that when applied, increased the intensity of images taken. So that our measurement closely aligns with the operating paradigms used for imaging short-lived phenomena, the measurements shown in this paper were achieved using a gate width of 1 μs and a gain of 4095. There has been much research into optimizing many different types of polarimeters [3,27–30]. Our decision to mount LCVRs at 0° and 45° for both the PSA and PSG was due to the ease of use and flexibility afforded by this setup. Using this equipment, mounted in this particular manner, meant that the PSG and PSA were capable of producing every Stokes vector on the surface of the Poincaré sphere [31], therefore producing an incredibly flexible system with nearly complete functionality. Furthermore, this alignment allowed for very robust control over the Stokes vectors, as the relation between each element of the Stokes vector and the retardances of the two LCVRs is very simple. An equation detailing this relationship is shown below. However, it should be clarified that at a working wavelength of 625 nm, our particular LCVRs were

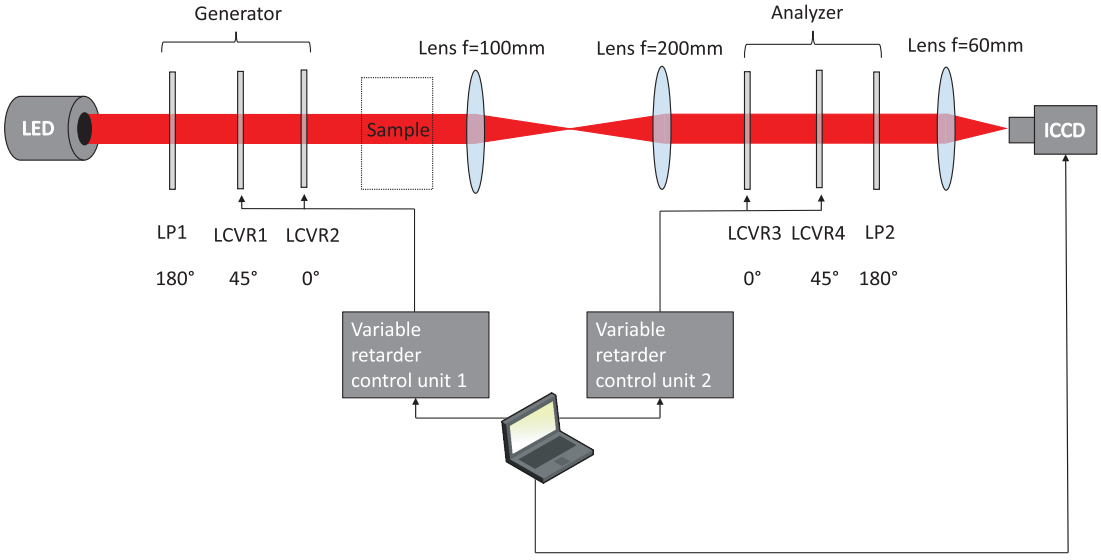


Fig. 1. Experimental setup consisting of a polychromatic red light LED, ICCD camera, and a sequence of linear polarizers, liquid crystal variable retarders, and lenses.

limited to a range of retardances of roughly 15°–350°, so the full Poincaré sphere was not quite accessible:

$$S = \left[\, 1,\ \cos(\delta_1)\ ,\ \sin(\delta_1)\cos(\delta_2)\ ,\ \sin(\delta_1)\sin(\delta_2)\,\right], \quad \textbf{(1)}$$

where δ_1 and δ_2 are the retardances of LCVR1 and LCVR2, respectively. It should be pointed out that this relation is the exact description of the x, y, and z coordinates in a spherical polar system, which greatly helps in understanding what effect each LCVR has on the ensuing Stokes vector. Two samples were used in this experiment, all of which are non-depolarizing: a quarter-wave plate (QWP) mounted at 20° and a compound sample of two plates of 0.5 mm thick bismuth silicon oxide (BSO) crystals [32]. The majority of the following results originate from the QWP measurements; the reason for this is that the Mueller matrix of a QWP is very well defined, and thus makes an excellent choice of sample for investigating the accuracy of results. Despite being a less than ideal system, BSO was also investigated because of its usefulness in the dynamic, low-exposure imaging settings that we were trying to recreate, and as such, it produces a system closer to those used in real-world dynamic polarimetry measurements. Eight Stokes vectors were chosen to be generated by both the PSG and PSA, generating 64 measurements in an 8×8 intensity matrix. To fully calculate a Mueller matrix, a 4×4 system using only 16 measurements are necessary, so with our 8×8 collection of 64 measurements, we can generate a large number of unique subsets with more than 16 measurements in them. In fact, using 64 measurements in the way, we have in this paper produced over 20,000 unique and valid intensity matrices of smaller dimensions, e.g., 5×6 or 7×7 (see Table 1), each producing a different Mueller matrix of its own, which we feel is a large enough collection of calculated Mueller matrices to draw conclusions upon. As we were using eight Stokes vectors, a good set to choose for this experiment is from the cubic set, a member of the vector sets known as Platonic solids [27,28]. The cubic set is so called because the vectors form the vertices of a cube within the Poincaré sphere [6,19]. The appeal of the cubic set is because it achieves the minimum condition number for the A and W matrices [33]:

$$A = \begin{bmatrix} 1 & -a & a & -a \\ 1 & -a & -a & a \\ 1 & -a & -a & -a \\ 1 & -a & a & a \\ 1 & a & a & -a \\ 1 & a & -a & a \\ 1 & a & -a & -a \\ 1 & a & a & a \end{bmatrix}, \quad W = A^T, \quad \textbf{(2)}$$

where A and W are the collection of Stokes vectors from the PSG and PSA, respectively, $a = \frac{1}{\sqrt{3}}$, and T denotes the matrix transpose. The Stokes vectors in the PSG and PSA were created by applying synchronized square waves of 1, 2, 4, and 8 Hz to the LCVRs. However, the 1 Hz square wave was placed 180° out phase with the others, and most crucially, the minimum and maximum voltages were the same, in effect creating a constant voltage. The reason for this choice is because it was used to trigger the camera, as the trigger would otherwise occur during the transition between one voltage to another. It was deemed best to keep the voltage constant to ensure stable retardance during image acquisition. Using this procedure meant that the PSA alternated between two states, and for each of these states, the PSG cycled through four, creating a sequence of eight measurements, and so producing the full 64 measurements in an additive manner. Square waves were chosen as they provide a greater degree of stability and consistency over continuously varying waveforms. Alongside this, the fact that the trigger was 180° out of phase with the varying retarders meant that they will have had enough time to stabilize after their transition. Because of inevitable small misalignments in the fast axes of LCVRs and the manner in which the Stokes vectors were created, it meant that it was not feasible to perfectly reproduce the values in the ideal A and W matrices. However, it was possible to reproduce the switching nature of the polarity of each element of the Stokes vectors, i.e., switching s_2 from $-a$ to a. In this way, each created Stokes vector occupied a different octant of the Poincaré sphere all to itself. To calculate the A and W matrices, first we calculated what the values of A were by measuring each vector against a known analyzer made of a QWP and a LP. Each vector was measured six times, each time against a different member of the diamond set of Stokes vectors. Once A was known, we then replaced the known analyzer with the remaining LCVRs, and then used these to measure A again. With this set of measurements, W could be calculated. Finally, we validated the matrices by comparing ideal and calculated values of the Mueller matrix of a QWP placed between the PSG and PSA, and with its fast axis mounted at 20°. The values calculated are shown in Eq. (3):

$$A = \begin{bmatrix} 0.992 & -0.849 & -0.062 & 0.395 \\ 0.981 & -0.855 & -0.338 & -0.059 \\ 0.998 & -0.878 & 0.095 & -0.220 \\ 0.995 & -0.866 & 0.212 & 0.206 \\ 0.981 & 0.437 & 0.462 & -0.729 \\ 1 & 0.544 & 0.663 & 0.536 \\ 0.966 & 0.587 & -0.532 & 0.542 \\ 0.961 & 0.516 & -0.569 & -0.596 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.978 & -0.810 & -0.438 & 0.260 \\ 0.973 & -0.753 & -0.320 & -0.510 \\ 0.986 & -0.816 & 0.402 & -0.386 \\ 1 & -0.858 & 0.316 & 0.308 \\ 0.978 & 0.323 & 0.575 & -0.711 \\ 0.972 & 0.193 & 0.669 & 0.662 \\ 0.958 & 0.282 & -0.612 & 0.672 \\ 0.956 & 0.373 & -0.644 & -0.584 \end{bmatrix}^T. \quad \textbf{(3)}$$

3. ASSESSMENT OF ACCURACY

The analyses involved in this paper are primarily concerned with the accuracy and stability of the results produced by different measurement schemes. To investigate these characteristics, several terms were introduced to act as proxies. The first of these terms is D, which we use as a surrogate measurement of accuracy; it is defined as the root mean square difference between the ideal Mueller matrix and the measured Mueller matrix:

$$D = \sqrt{\frac{1}{16}\sum_{i,j=1}^{4}\left(M_{i,j}^{ca} - M_{i,j}^{id}\right)^2}, \quad \textbf{(4)}$$

where $M^{ca}_{i,j}$ and $M^{id}_{i,j}$ are the Mueller matrix elements of the calculated and ideal matrices, respectively. To understand how the stability of results changes, an indirect yet relevant measurement that evaluates the standard deviation of the Mueller matrix across each pixel was developed, and we call this value ρ (for examples of this, see Figs. 4 and 5):

$$\rho = \frac{1}{16}\sum_{i,j}^{4}\sqrt{\frac{1}{1024^2}\sum_{n=1}^{1024^2}(M^{n}_{i,j} - \overline{M_{i,j}})^2}, \tag{5}$$

where n indexes the pixel, i and j index the particular element in the Mueller matrix, and $\overline{M_{i,j}}$ is the average value of the Mueller matrix element across the whole image. To simplify the figures in the Results section and aid in readability, we define the dimension of the A and W matrices as N_A and N_W, respectively. Essentially, what these numbers denote are the numbers of Stokes vectors produced in the PSG and PSA. For example, the scheme $N_A = 6$ and $N_W = 5$ means that six Stokes vectors are produced in the PSA and five Stokes vectors are produced in the PSG, i.e., the dimensions of the A and W matrices are (6×4) and (4×5), respectively. To probe how systematic error manifests in our results, we also introduce a new term that we call the Δ matrix. Its definition and derivation are outlined as

$$\Delta = I_c - I_m, \tag{6}$$

where I_c is the matrix containing the intensity measurements derived from the calculated Mueller matrix, M_c, and I_m is the matrix containing the intensity measurements from the camera:

$$I_m = A_{ac} M W_{ac}, \tag{7}$$

$$M_c = A_{as}^{-1} I_m W_{as}^{-1} \quad M_c = A_{as}^{-1} A_{ac} M W_{ac} W_{as}^{-1}, \tag{8}$$

where the lower scripts ac and as denote the actual and assumed versions, respectively. What is meant by actual and assumed is that the actual Stokes vectors are those used directly in the measurement procedure and the assumed vectors are those used in the calculation of the Mueller matrix, such that in a perfectly calibrated system, we would have $A_{ac} = A_{as}$ and $W_{ac} = W_{as}$. It should also be pointed out that by definition, we cannot know A_{ac} and W_{ac}, and therefore we cannot know M itself. We do of course know I_m, and that is why the Δ matrix is defined using I_m. This way, the delta matrix can be calculated even if the Mueller matrix is not known. This is in contrast to our parameter D, which works only when using a reference sample:

$$I_c = A_{as} M_c W_{as}, \quad I_c = A_{as} I_m - I_c = A_{ac} M W_{ac} - A_{as} M_c W_{as}. \tag{9}$$

From this definition of Δ, we can see that it originates from the statistical noise in I_m and the difference between the actual Stokes vectors and their assumed counterparts used for calculation, i.e., systematic error. Of course, one can calibrate away a large degree of systematic error so that A_{as} and W_{as} are very close to their actual counterparts. However, it is impossible to eliminate it entirely, and in an over-determined system, many Stokes vectors are used, so it is important that we have a method for identifying systematic error. This divergence between the intensities is an incredibly curious and seemingly redundant result, and indeed in a perfect system devoid of error, it does not exist. It is a consequence of both over-determination and the existence of measurement errors. It is important to note that it is not a sole result of over-determination, and that an over-determined system free of error will have an empty Δ matrix. The reason for its existence is that the statistical errors in the measured intensities and the systematic error in the A and W matrices create a set of logically inconsistent linear equations, i.e., there is no Mueller matrix that can perfectly satisfy the relation $I = AMW$. To illustrate this point, see the following example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 6.2 \end{bmatrix}. \tag{10}$$

As you can see, this equation contains a mathematical inconsistency. It is impossible to have a vector whose elements are one, two, and three, but the sum of those elements is equal to 6.2, and yet this equation is still solvable (see Appendix A) *Invertinga Rectangular Matrix*. It is this possibility of solving logically contradictory equations that gives rise to the Δ matrix. We believe that the magnitudes in the Δ matrix can be thought of as the degree of contradiction between the linear relations collected in the equation $I = AMW$. It is important to reiterate that when one over-determines a Mueller matrix, an unnecessarily large set of measurements is taken, and with this overly large set, smaller subsets can be assembled by omitting rows and columns. In turn, some of these subsets can be used to produce their own Mueller matrices (which is what we mean when we refer to a "result"). We hasten to add that over-determination does not automatically produce multiple results; it simply means that it is now possible to produce multiple results. So this means that we have gone from dealing with a single result to sets of results collected together by their shared values of N_A and N_W (i.e., the Mueller matrices produced by all possible $N_A = 4$ and $N_W = 4$ combinations is a collected set of results). The size of each set of results (which we denote as N) is dependent upon N_A, N_W, and the maximum of N_A and N_W, which for this paper is eight. The size of each set is defined in Eq. (11), which is the product of combinations for a given N_A and N_W:

$$N(N_A, N_W) = \left(\frac{8!}{N_A!(8-N_A)!}\right)\left(\frac{8!}{N_W!(8-N_W)!}\right). \tag{11}$$

The associated values of N for all values of N_A and N_W used in this paper can be seen in Table 1. As one can see, the size of the sets can grow extremely rapidly. Due to the inherent multiplicity of this methodology, the following results will involve the averages and minimums across these sets. In this paper, we

Table 1. N, Total Number of Possible Combinations for a Chosen N_A and N_W

N	4	5	6	7	8
4	4900	3920	1960	560	70
5	3920	3136	1568	448	56
6	1960	1568	784	224	28
7	560	448	224	64	8
8	70	56	28	8	1

develop two simple terms to characterize accuracy and stability, namely, D and ρ, respectively. Although useful, these two terms do not provide the full picture. Their limitation arises from their reduced scope; they relate only to the results, not to the matrices that produced them. To gain a better understanding of the dynamics between the A and W matrices, and the quality of results they produce, we calculate two features of these matrices: the condition number (C) and the equally weighted variance (EWV). It has been shown previously [3,6,19,21,33] that minimizing these two terms for both A and W matrices will result in optimal measurements (a relationship we probe in Fig. 3). The calculation of these terms is not too difficult [see Eqs. (12) and (13)]; one issue though is that they can both very easily go to infinity. As such, we have shown the Log_{10} of each of these values, so that the full range of values can be compared:

$$\kappa_A = \|A\|\|A^{-1}\|, \tag{12}$$

where $\|\cdot\cdot\|$ denotes the Frobenius norm. It should be pointed out that the calculation of the condition number is a general procedure and can be performed for any matrix (provided it is non-singular); we have simply chosen to show the condition number of the A matrix (κ_A) for this example:

$$EWV = tr[(A^T A)^{-1}], \tag{13}$$

where tr denotes the trace of the matrix.

A. Simulating Noise

Simulations were conducted in parallel with the physical experiments, allowing for the verification of our findings. Simulating these experiments was no great task, as a piece of simple code can easily run the matrix algebra necessary to perform the calculations required. However, the difficulty arises in the modeling of the noise. For our simulations, the noise was split into statistical noise and systematic error. The statistical noise was intended to represent the fluctuations in the light source and the camera, resulting in fluctuations in the number of counts the camera detects. The systematic error acts as a substitute to the error caused by differences in the Stokes vectors used for measurement and those used in calculation, i.e., one believes the PSG was generating the vector [1,1,0,0] when making measurements, so this is what is used when calculating the Mueller matrix, but in reality, the PSG was actually generating [1,0.979,0.199,0.044]. To model the statistical noise, we used additive Gaussian noise with a distribution centered at zero and variance ranging from 0% to 25% of the value of m_{00}, the transmittivity of the Mueller matrix, where the magnitude of the standard deviation of this distribution, relative to the maximum possible intensity, represents the statistical noise. Modeling the systematic error was a little more involved. In the previous section, we introduced the notion of actual and assumed versions of A and W, i.e., the real matrices used to create measurements and the assumed matrices used to calculate the Mueller matrix. To simulate this framework, the matrices were split as outlined, where the actual versions contain the assumed Stokes vectors plus the systematic error:

$$A_{as} = \begin{bmatrix} 1 & \cos(\phi_0) & \sin(\phi_0)\cos(\theta_0) & \sin(\phi_0)\sin(\theta_0) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(\phi_n) & \sin(\phi_n)\cos(\theta_n) & \sin(\phi_n)\sin(\theta_n) \end{bmatrix}, \tag{14}$$

$$A_{ac} = \begin{bmatrix} 1 & \cos(\phi_0+\delta) & \sin(\phi_0+\delta)\cos(\theta_0+\delta) & \sin(\phi_0+\delta)\sin(\theta_0+\delta) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(\phi_n+\delta) & \sin(\phi_n+\delta)\cos(\theta_n+\delta) & \sin(\phi_n+\delta)\sin(\theta_n+\delta) \end{bmatrix}, \tag{15}$$

where δ represents the systematic error and is measured in degrees.

4. RESULTS

A. Accuracy and Stability

In this section, we show that over-determination has a positive effect on both the accuracy and stability of the Mueller matrix produced; we also explain some of the more complex relationships between N_A and N_W and the most accurate results produced. We begin by showing an analysis of D, the difference between the actual and calculated Mueller matrix. In Fig. 2, the average of D taken across the sets created by different values of N_A and N_W is displayed, and it clearly shows that on average, increasing N_A or N_W results in a calculated Mueller matrix that is closer to the actual matrix. The largest gains can be made by increasing either N_A or N_W from four, but there are diminishing returns as N_A or N_W is increased, so much so that increasing N_A or N_W beyond a certain point will not likely provide any noticeable improvement. Furthermore, the consistency and symmetry of the change in $\overline{D}$ should also be noted; what this means is that increasing N_A has the same effect as increasing N_W. Two very clear reasons for the behavior illustrated in Fig. 2 are shown in Fig. 3. Here we see that increasing the size of both the A and W matrices results in simultaneously decreasing their EWV and condition number. Reductions in both of these matrix characteristics have been linked to enhanced performance in Stokes polarimetry [6]. Figure 3 easily demonstrates that this relationship still holds in Mueller polarimetry. Figure 4 shows the results from the investigation into the impact N_A and N_W have on ρ, the spatial standard deviation of the calculated Mueller matrix. The most noteworthy aspect of this figure is that an experimental scheme using either $N_A = 4$ or $N_W = 4$ can produce extremely unstable results. The cause of the divergence between experimental and simulated results for $N_A = 4$ is not clear, but one possible reason is that the sample sizes are not large enough relative to the standard deviation, so the averages have not stabilized. Another possibility is that the collection of $N_A = 4$ results contains many results from poorly conditioned and highly redundant A matrices, which are highly susceptible to systematic error. This susceptibility, alongside larger numbers of results (see Table 1), increases the likelihood of several extremely high values of ρ, which could potentially skew the averages shown in Fig. 4. It is also interesting to observe that once N_A and N_W exceed four, the relationship between

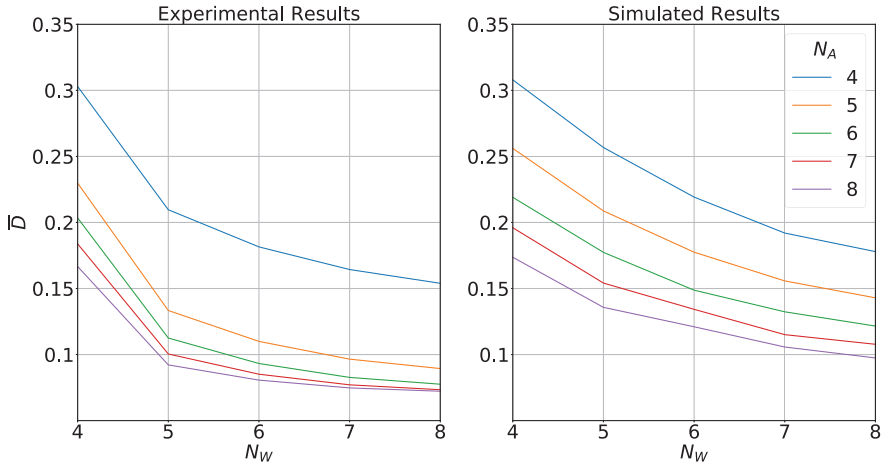


Fig. 2. Experimental and simulated QWP results showing the average difference between calculated and ideal matrix ($\overline{D}$), taken across the sets created by different values of N_A and N_W.

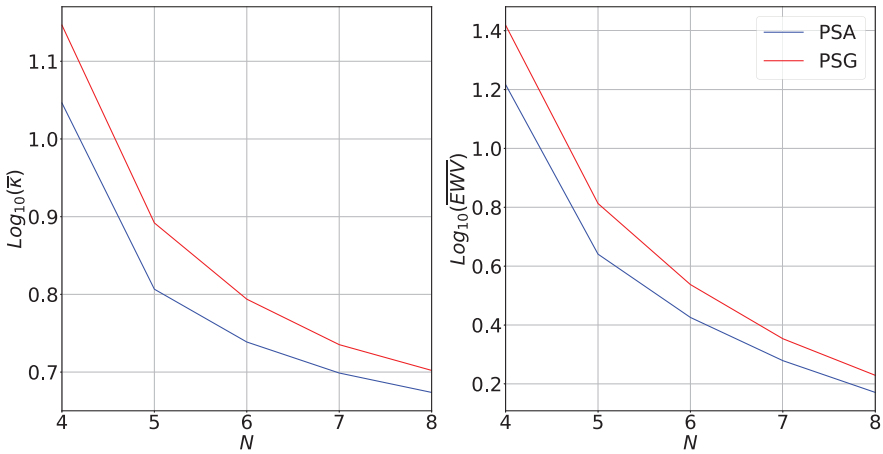


Fig. 3. Experimental QWP results showing the Log_{10} average equally weighted variance and condition number for A and W matrices, denoted by PSG and PSA, respectively. Each average was taken across the sets created by different values of N_A and N_W.

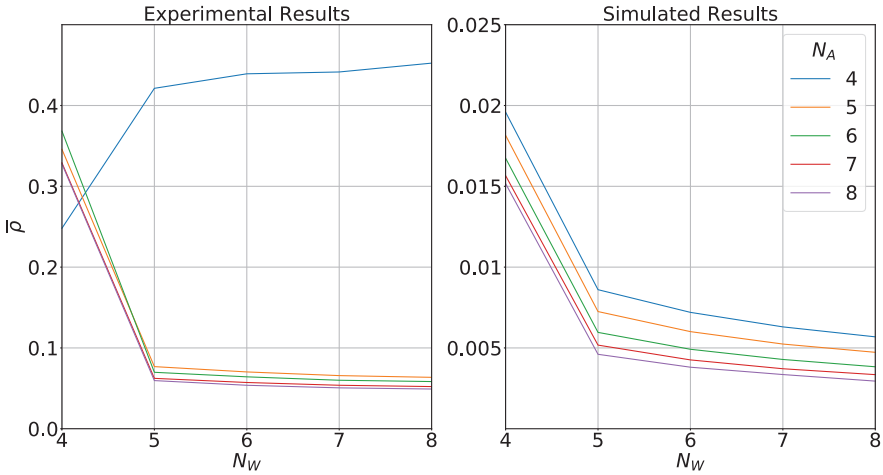
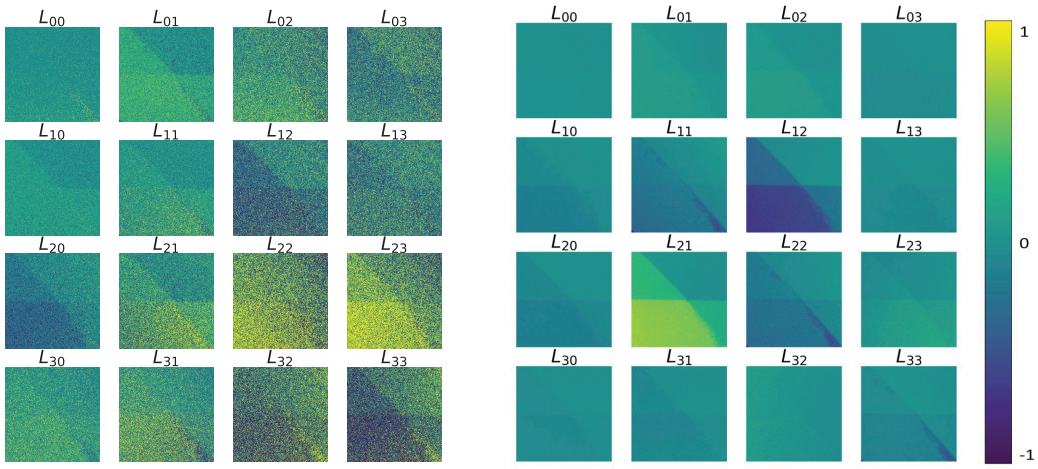


Fig. 4. Experimental and simulated QWP results showing the average ρ against different values of N_A and N_W.

$\overline{\rho}$ and N_A and N_W appears to be linear, which is in contrast to Fig. 2, which shows that increasing N_A or N_W has a nonlinear effect on $\overline{D}$. An accompaniment to Fig. 4 is provided in Fig. 5, where a much more tangible interpretation of ρ is illustrated. Both figures show the matrix logarithm of the [34–36] Mueller matrix of a compound sample of air and two separate, partially



(a) Logarithm of Mueller matrix using a 4x4 measurement system

(b) Logarithm of Mueller matrix using an 8x8 measurement system

Fig. 5. Stability comparison between a 4×4 and 8×8 system on a compound sample of two partially overlapping BSO crystals. The color bar applies to both plots.

overlapping 0.5 mm thick crystals of BSO, where the matrix logarithm is simply $L = Ln(M)$ [37]. If we look at the L_{21} sub-image on the right-hand side of Fig. 5, we can clearly see that the image is split into four distinct areas, each corresponding to a different element of the compound sample. Starting from the top right corner of the image and moving clockwise, the sample element is air, a non-overlapped 0.5 mm thick BSO crystal, an overlapped compound of 1.0 mm thick BSO, and a non-overlapped 0.5 mm thick BSO crystal. The difference between the two figures is that the left image was produced using a 4×4 measurement scheme (the chosen indices of the A and W matrices were [0, 1, 6, 3] and [0, 1, 2, 6], respectively), whereas the right used an 8×8 scheme. It is immediately apparent how consistent the results are between the pixels in the 8×8 system compared to the 4×4. This highlights just how incredibly susceptible a 4×4 system is to statistical noise; just a small, inevitable change in the value measured between pixels in the intensity matrix can result in a large change in the ensuing Mueller matrix, leading to the blurriness in Fig. 5(a). Nevertheless, the results produced by the 8×8 system are not perfect. As stated previously, the sample consists of two partially overlapping pieces of BSO, which is a non-depolarizing medium [38–40]. What non-depolarizing means in terms of the Mueller matrix, and the matrix logarithm, is that the values in the first row and column should be identical, and the off-diagonal elements should be anti-symmetric with each other, i.e., $M_{12} = -M_{21}$. As Fig. 5(b) demonstrates, the values in the first row and column are small but certainly non-zero. However, it is not expected that the samples become perfectly non-depolarizing, as there are bound to be imperfections introduced in production, shipping, handling, etc. It should also be highlighted that the logarithm of the Mueller matrix was shown specifically to highlight the additive nature of the Mueller matrices where they overlap. As both crystals have the same cut (0,0,1), thickness, and orientation, they should have identical Mueller matrices; therefore, the Mueller matrix of the section where they overlap should be M^2, and thus L in this section should be exactly double the value shown in the single layer

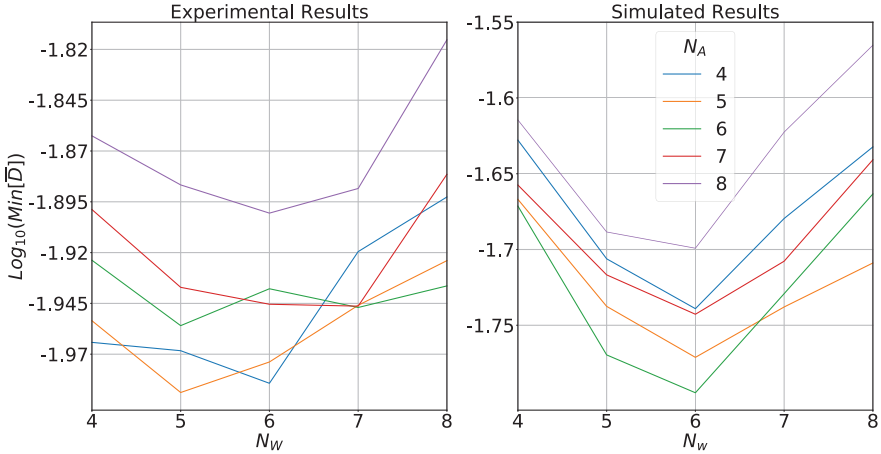
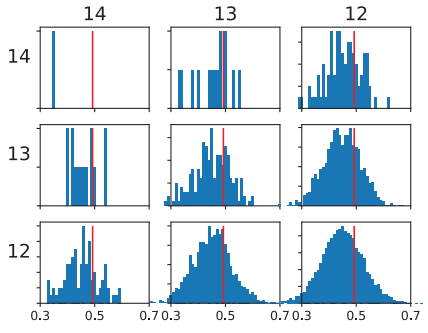
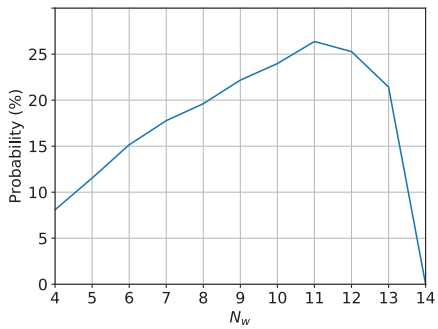


Fig. 6. The Log_{10} minimum of experimental results of a double layer of BSO compared with simulated results (see Dataset 1, Ref. [44]).



(a) Distributions of calculated results of an element within a Mueller matrix Vs N_A and N_W. The vertical red line represents the true value.



(b) Probability of a Mueller matrix element being within 2.5% true value for N_A = 14

Fig. 7. Opposing effects of increased accuracy versus decreased set size, which are crucial for understanding the dynamics of over-determining a Mueller matrix (see Dataset 2, Ref. [45]).

sections. This can be clearly seen in L_{12}, where the single layer sections have roughly equal values of -0.2, but the double layer section has a value of -0.4. As stated previously, over-determining a Mueller matrix produces sets of results rather than a single result. Therefore, when performing an analysis of these results, one can look at the average values across these sets, or at the minimum and maximum values appearing in these sets. In Fig. 6, we can see the results of such an analysis, where we show the log of the minimum value of D calculated in each set of results produced by different values of N_A and N_W. When compared to Fig. 2, this figure seems to contradict the result that increasing N_A and N_W gives a lower D, however Fig. 2 was showing the average across each set, whereas this analysis is investigating the minimum. The figure shows that for each N_W, increasing the N_A value from four will produce a set of results that contains a Mueller matrix that has a lower D value than that in the preceding set. Nevertheless, once a certain value of N_A has been exceeded, this effect will reverse, and increasing N_A will produce a set of results that has a larger minimum value of D in it. It should also be noted that the experimental results in Fig. 6 overlap each other far more than the simulated results. We believe one potential cause is due to two reasons: first, there is a potential for overlap in the condition number and EWV of similar result sets, for example, some of the combinations in the $N_A = 5$ and $N_W = 6$ sets will have condition numbers very similar to combinations in the $N_A = 6$ and $N_W = 6$ sets. Second, the higher levels of statistical noise in the experimental results will increase the variance of the results, and thus make overlaps more likely. Overall, Fig. 6 shows that calculating a Mueller matrix using more measurements does not necessarily produce a better result; in fact, it can produce a worse result. To explain this unexpected conclusion, it must be understood that there are two effects at work in Fig. 6. First, there is the beneficial effect (shown in Fig. 2) of increasing N_A and N_W, producing sets that, on average, are increasingly accurate; however, there is a second retarding effect working against this positive relationship. This effect is the negative relation between the size of the set and the magnitude of N_A and N_W. This point can be seen in Table 1. The size of the set produced by the 4×4 system is 4900, but the size of the 8×8 system is only one. The interaction of these two opposing effects can be seen in Figs. 7(a) and 7(b). Figure 7(a) shows a continuation of the simulations conducted to make Figs. 2 and 4. Here we see the distributions of calculated values of a single element of the Mueller matrix for different values of N_A and N_W, where the red line denotes the true value, the x axis is the binned values, and the y axis is the number of results in each bin. This figure was developed using simulations so that larger values of N_A and N_W could be used, and thus produce larger sets to aid visualization. To do this, we extended our previous choice of eight Stokes vectors from the cubic set of vectors, by including the six vectors from the diamond set of vectors, i.e., $[1, a, a, a]$, $[1, -a, a, a]$, etc., for the cube, and $[1, 1, 0, 0]$, $[1, 0, 1, 0]$, etc., for the diamond [28]. As the figure shows, reducing N_A or N_W increases the size of the set, producing a Gaussian distribution centered close to the true value. As the set size increases, the number of results far from the true value also increases, so that if we now imagine a situation where we want to pick a single result from each set that is as close to the true value as possible, larger sets would have a lower probability. This situation is shown in Fig. 7(b), where the probability of the calculated value of each set being within 2.5% of the true value is shown. As it shows, increasing N_A from four has a positive effect on the probability, as the average value of each set gets closer to the true value. However, we see that this effect stops at $N_A = 11$ and goes into reverse, until finally the probability goes to zero at $N_A = 14$. This is because at this point, there is only one result to choose from in this set, and it is outside the 2.5% range. This chart clearly explains the two effects of increased accuracy versus reduced set size occurring in Fig. 6.

B. Quantifying Noise

In this section, we will show how the various metadata terms that arise through the over-determination process can be used to quantitatively approximate the statistical noise and systematic error within the measurements. First, we begin with the Δ matrix, which we introduced in the assessment of accuracy section and defined as an artifact of logical fallacies within the relation $I = AMW$. In this section, we will show the dynamics of this term and how it can be used to benefit the experimentalist. To create Fig. 8, simulations were conducted where the statistical noise and systematic error were varied, then the root

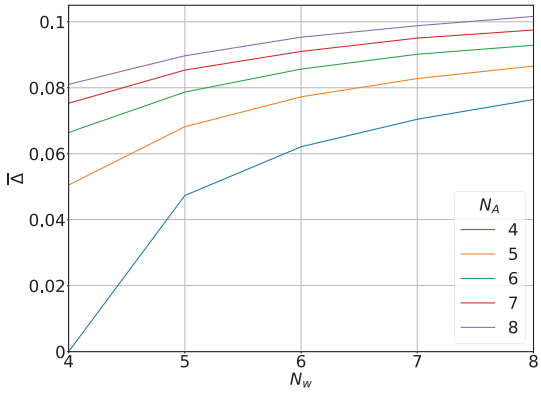


Fig. 8. Average magnitude of elements of Δ taken across multiple noisy systems versus N_A and N_W.

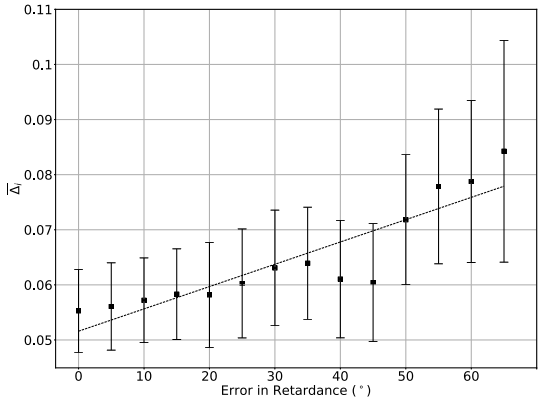


Fig. 9. Row average and standard deviation of Δ versus systematic noise associated with each row.

mean square of the elements of Δ were calculated for all measurement schemes. The systematic error was varied as shown in Eq. (15), with values of δ ranging from 0°–60°. Statistical noise was modeled using an additive Gaussian distribution centered on zero, with a variance ranging from 0%–25% of m_{00}. The values from the root mean square of the elements of Δ were collected together in their respective measurement schemes and averaged, creating an average taken across many different levels of noise. In this figure, we can see that the magnitude of the elements of (Δ) increases with respect to both N_A and N_W, which is interesting given that the average value is displayed, rather than the sum, which would of course increase with N_A and N_W, as the dimensions of Δ are (N_A, N_W). So what this means is that despite having equal levels of statistical and systematic error, increasing either N_A or N_W will still increase the average magnitude of terms within Δ, i.e., in a noisy system, increasing N_A or N_W makes it increasingly difficult for a Mueller matrix to perfectly satisfy $I = AMW$. Perhaps the most interesting artifact of this figure is that it shows that regardless of the level of noise, a 4 × 4 system will create a Mueller matrix that perfectly satisfies $I = AMW$, and thus produces an empty Δ. The reason for this is because the relations between measurements collected in matrix equations can become mathematically inconsistent only once there are more than enough measurements to prove that it is so. For example, in the case where only 15 measurements have been taken of an unknown Mueller matrix, we cannot say if the results are inconsistent with each other, as there are in fact an infinite number of Mueller matrices that could perfectly satisfy $I = AMW$. In the case of a 4 × 4 system, we have reached the information threshold and can start proving if new measurements are mathematically consistent with the existing ones; however, if there are inconsistencies in this system, there will be no solution at all. Therefore, until the Mueller matrix is over-determined, there can be no mathematical inconsistencies like those shown in Eq. (10), even if un-physical values are used in the calculation. One of the difficulties of Δ is that the noise of each measurement becomes entangled and distributed across all terms. For example, a 5 × 5 system with 24 noiseless and one noisy measurement will produce a Δ with non-zero elements that correspond to the noiseless measurements. Nevertheless, it is still possible to disentangle some of the information represented in Δ to provide a quantitative approximation of the systematic error associated with each Stokes vector. As Δ follows the same structure as I, rows correspond to a single vector from A, and columns correspond to a single vector from W. We can exploit this by examining the sum of each row and column within Δ to ascertain the level of noise associated with each

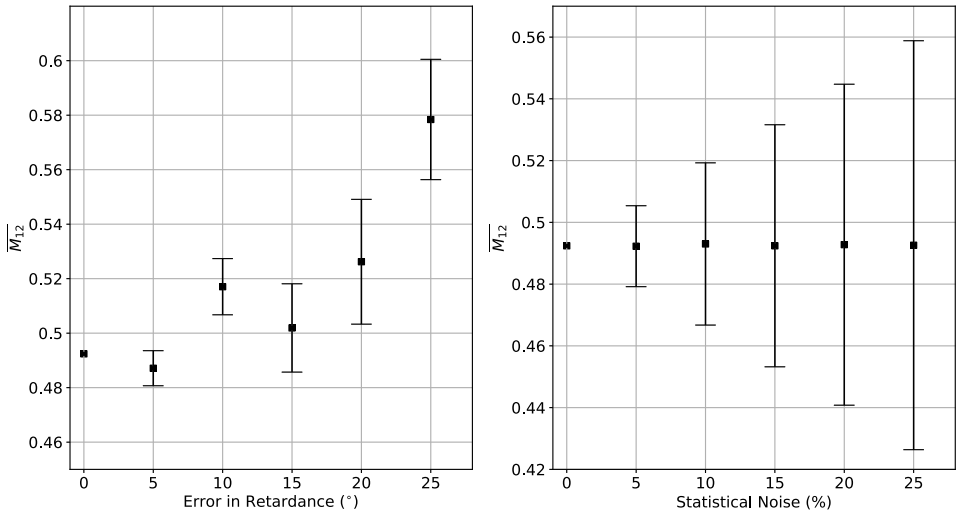


Fig. 10. Comparison between the effects that statistical and systematic errors have on the distributions of calculated values of the Mueller matrix.

vector. To perform this investigation, simulations were once again conducted, where this time the statistical noise was kept constant and the systematic error for the PSA was zero, but the error in retardance was varied for each vector from the PSG, so that each of the 14 vectors had a different error associated with it. Then the Δ matrix was determined, and the sum of the absolute values of each row was calculated and plotted against the associated vector's error in retardance. The results of this can be seen in Fig. 9. This figure clearly demonstrates that raising the error in retardance for a given PSG vector, increases the magnitude of elements within Δ associated with that vector. This result paves the way for a very simple identification and quantification of systematic error associated with each vector used in the measurements (for PSA vectors, the column sum is taken instead of the row sum). As stated previously though, the information contained within Δ is heavily entangled and also contains statistical noise as well as systematic. Therefore, this technique can be used only as a relative indication of the systematic error associated with each vector, rather than an unequivocal direct calculation. It is best used in a relative sense; for example, if eight of the vectors used have values around 0.02, yet one has a value of 0.06, then it can be safely assumed that this particular vector has a large amount of systematic error associated with it. Now that we have outlined a system for identifying systematic error, we will continue looking at distributions of results of a single element of the Mueller matrix, except now we will focus on what effects systematic and statistical noise has on the shape of distributions shown in Fig. 7(a). For this investigation, we will look at the average and the standard deviation of these distributions under different levels of systematic and statistical noise. To perform this analysis, we conducted another simulation where systematic and statistical noise was eliminated but the other noise was kept, i.e., no statistical noise with systematic error and vice versa. The results of this investigation can be seen in Fig. 10. To simplify the results, we have narrowed down the list of elements Each chart in Fig. 10 demonstrates a simple relationship between each type of noise. Starting with the systematic error, we can see that increasing this type of noise primarily results in a shift in the average of the distribution; it is also accompanied by a small increase in the standard deviation. If we turn our attention to the statistical noise on the other hand, we see that the average is unchanged; instead, there is a very large increase in the standard deviation, an increase that is far larger than that of the systematic error. Although these two contrasting relationships are clear, we hasten to point out that these two types of noise are not easily equatable, so it is difficult to say with utmost confidence that statistical noise has a much larger effect on standard deviation, when compared to systematic error. It is rather fascinating to see that the shift in average due to systematic error is positive in this particular example. It would be interesting to see if this shift is always positive, or if it varies depending on the Mueller matrix being measured, the choice of Stokes vectors, or a bias in the noise itself; sadly, these questions are beyond the scope of this paper.

5. DISCUSSION

Two of the largest benefits of over-determining a Mueller matrix are that multiple results are produced, and that these results are, on average, increasingly accurate and stable. However, there is still the question of what degree of over-determination is necessary. The answer to this will depend upon the measurement equipment, the Mueller matrix to be measured, and the desired accuracy. For example, a complex depolarizing Mueller matrix is to be measured with equipment with relatively small statistical noise, so what measurement scheme should be used? To answer this question, it is important to note that when one uses an 8×8 measurement scheme, the reason to do so is not necessarily to gain access to the single result using all 64 measurements, because as we have shown, this one result will likely be worse than many of the results from a lesser over-determined system. Instead, the purpose is to generate large sets of 6×6, 6×5, 5×5, etc., results, and from these sets, one is much more likely to find an accurate and stable result. So even if a 4×4 scheme is deemed satisfactory, it is still advantageous to increase the measurement scheme so that there is a larger set of 4×4 results to choose from. However, according to our results, heavy over-determination such as an 8×8 measurement scheme is only slightly more beneficial than a moderately over-determined, such as a 6×6 scheme, so despite believing that any over-determination will always be superior to a simple, well-determined system, we believe that there is not much benefit in using a heavily over-determined scheme. The expected behavior of the Δ matrix would be that the average magnitude of the terms within Δ would be dependent only on the degree of noise in the system and completely independent of the measurement scheme. However, we have shown that increasing the degree of over-determination in a noisy system will make finding a Mueller matrix that closely satisfies $I = AMW$ less likely. One possible factor behind this behavior is the way that the errors are entangled and distributed among the terms. If each element within Δ contains information related to every measurement, then as the number of measurements increases, each term will contain more information, and thus on average, have a larger magnitude. However, this clearly does not explain the full behavior. If this were the only factor at work, then the average magnitude would be directly correlated to N_A and N_W, and would go to infinity alongside either parameter, but as Fig. 8 shows, this does not happen. Instead, the increase in average magnitude becomes successively smaller. Clearly there must be another factor working against this. One potential clue is shown in Fig. 2. As stated previously, increasing N_A or N_W produces a more accurate result on average, but for each successive increase in N_A or N_W, the increase in accuracy is reduced. One way of interpreting this behavior is that with each progressive measurement taken, the effect this new information has on the resulting Mueller matrix is reduced, i.e., the difference between Mueller matrices calculated in a 4×4 and 4×5 system is much larger than that between a 10×10 and a 10×11. Applying this principle back to Δ, we can see how the decreasing difference between calculated Mueller matrices when N_A and N_W increase can lead to a consequent diminution of delta as a function of N_A and N_W. Overall, we state that the total error is entangled and distributed among the elements of Δ, causing the magnitude of these terms to increase with N_A and N_W. Then we have the retarding factor of increasingly static calculated Mueller matrices that reduces the dependence of Δ with N_A and N_W. Now that we have a rationale of the behavior of Δ, we can move onto the second strange phenomena: why do only 4×4 measurement schemes have an empty Δ matrix? To understand this,

we must first understand how each set of four vectors from the PSG and PSA impacts the calculated Mueller matrix, beginning with the PSA. We start by stating that each measurement in I is the dot product of S', with a Stokes vector from the PSA. To simplify this explanation, we will switch from working with 4D to 3D Cartesian vectors. We begin with an unknown vector V_u that we would like to determine. To do this, we measure the dot product of this vector with three known vectors V_1, V_2, V_3. The first dot product, $V_u.V_1 = C_1$, is not very useful, as there is an infinite flat plane of possibilities in x, y, and z that satisfies this condition. A second dot product, $V_u.V_2 = C_2$, provides another infinite flat plane of solutions that intersects the first, creating a line where both dot products are satisfied. Finally, we perform the last dot product, $V_u.V_3 = C_3$, resulting in a new plane that will pass through this line, turning it into a single point, and it is this point that satisfies all three dot product equations, and thus V_u is now determined. Returning to the world of 4D Stokes vectors, we can now see that with each successive measurement of S' versus different vectors from the PSA, the range of possibilities is reduced from a 4D surface to a 3D plane then to a line and finally a point. So when four vectors are used to calculate S', they either intersect at a single point and produce one result, or they do not intersect enough, and thus S' cannot be determined. Now let us imagine that we measure S' against five vectors from the PSA. In a noiseless system, all five planes will intersect at a single point, removing the need to take a fifth measurement. However, if noise is introduced, there will no longer be a point where all planes intersect; instead, there will be five separate points where at least four of the planes intersect, and thus S' can be calculated, but Δ is now non-null. Applying this principle to the interaction between the Mueller matrix and the PSG, where it is best to think of the Mueller matrix as a collection of four column vectors, where each column has a separate effect, e.g., the first column can be thought of a vector that relates to S_0, the second to S_1, the third to S_2, etc. We can retain generality, and simplify this explanation by reducing the scope of the full Mueller matrix calculation to just the first column and assume that all four S' have been calculated using four PSA vectors. We repeat the process we applied to the PSA, except this time the unknown vector is the first column vector of the Mueller matrix. The first vector from the PSG and the associated S_0 value from S' describe a 4D surface; the second vector and calculated S_0 produce another 4D surface that intersects the first one, creating a 3D plane of solutions that satisfies both dot products. The following two vectors from the PSG reduce this 3D plane to a line and then finally a point, in the exact same manner as the PSA. This process can be repeated for every column, until the whole Mueller matrix is populated with column vectors that perfectly satisfy their respective dot products. It is this twofold process within the PSG and PSG of reducing the possible solutions to exactly one point that gives rise to the unerringly perfect satisfaction of $I = AMW$ within the 4×4 measurement scheme. If there are any mathematical inconsistencies within a 4×4 system, then this process of reducing the possibilities to a single point cannot happen, and thus the Mueller matrix cannot be calculated. Perhaps more interesting, it is likely this absence of a Δ matrix that leads to the poor performance of 4×4 schemes (especially when considering ρ). Because in a 4×4 scheme $I = AMW$ has to be perfectly satisfied, all the noise is directly present in the calculation, whereas in an over-determined system, the noise clearly does not always manifest itself; otherwise, there would be no Δ matrix at all. This leads us to believe that it is plausible that I_c in Eq. (6) is a less noisy equivalent of the measured intensities. One point that must be considered in this analysis is that both sets of Stokes vectors utilized (cubic and diamond sets) have the same absolute values of all their coefficients, i.e., a in the cubic and one in the diamond set. This is not the case for numerous other sets, and so it begs the question of what would happen to our results if different sets of vectors were used. A full investigation of this is out of the scope of this paper; however, we believe that the relationship shown in Fig. 3 is fairly general, insofar as the simplest way of reducing condition number and EWV is to add more vectors to the A and W matrices, even if the additional vectors are sub-optimal. It has also been shown many times that both EWV and condition number are directly linked to the accuracy of results [3,6,19,21,33]. By extrapolating from this fact, we believe that the relationships shown in Figs. 2 and 4 are also general and not a product of the Stokes vectors chosen. We are further validated in this belief by the results of a similar work [41] that uses Stokes vectors different from our experiments, and that details very similar relationships between performance and number of Stokes vectors used. Another potential area that requires further investigation is in understanding how the delta matrix behaves when different sets of Stokes vectors are used, particularly in the context of less-symmetric sets than those used in this study. Specifically, we currently have no evidence that either confirms or denies the claim that different Stokes vectors could produce different behavior in the delta matrix, especially when asymmetrical sets of Stokes vectors are used (e.g., the $N = 7$ set from the platonic solids). So we cannot say that the relationship shown in Fig. 9 is a general relationship that holds true for all Stoke vector sets.

6. CONCLUSION

In this paper, we have expanded upon previous research [3,20,21,33,41,42] into optimizing Stokes polarimetry by investigating the effect that over-determination has in the Mueller polarimetry regime. To perform this analysis, we conducted physical experiments alongside simulations, the combination of which not only highlighted relationships, but also allowed for logical explanations that provide a fuller understanding of the factors at work within the over-determination process. By investigating relationships between our terms D and ρ, and the degree of over-determination, we have found that over-determination is a worthwhile route to improving the accuracy and consistency of results derived from noisy images. It is hoped that this insight will allow Mueller polarimetry to enter into a wider range of disciplines, where previously high levels of noise presented a barrier to entry. Moreover, experimentalists already using Mueller polarimetry will be pleased to know that they can improve their results with this relatively simple technique. Alongside our inquiry into over-determination, we also conducted novel research into the particular effects that systematic and statistical noise has on Mueller polarimetry. There already exists much work focused on understanding and managing statistical noise; however, there is little work in the existing literature that scrutinizes systematic error—a point our paper addresses. In the Quantifying Noise section, we provide clear relationships between both types of noise and the

different effects they have on distributions of results, namely, that statistical noise simply increases the standard deviation without affecting the average, and that systematic error shifts the average and slightly increases the standard deviation. After demonstrating these effects, we go on to provide a method for their quantification, which in the case of systematic error is a breakthrough, as previously there was no means to determine the magnitude or the origin of this type of noise. To do this, we introduce a new term that we call the Δ matrix. We show that this term is directly linked to the systematic error within each Stokes vector used in the measurement process. This Δ matrix is therefore a very powerful tool, insofar as it allows the experimentalist to identify specific Stokes vectors that are responsible for the systematic error. Although we are pleased with the results of this paper, there is certainly scope for improvement. In particular, it would be interesting to see this investigation repeated for different measurement schemes involving all the platonic and non-platonic solids, rather than the cubic system chosen for this investigation [43]. It could also be worthwhile to replicate some of these investigations, especially Fig. 2, except measurements with large condition numbers and EWV are filtered out so that only the best of each measurement scheme is compared. Additionally, a more thorough theoretical investigation into the effects of systematic error should be conducted; it is highly likely that a more realistic method for simulating systematic error can, and should, be devised. Finally, we would like to see more work being done understanding the Δ matrix. It would be very interesting to see if the errors associated with each measurement could be thoroughly disentangled. The benefit of this is that it could potentially provide a route for the reduction of noise from measurements—a prospect that would be greatly beneficial for those within and outside the Mueller polarimetry community.

APPENDIX A

1. CALCULATING AN OVER-DETERMINED MUELLER MATRIX

In this section, we detail how to use the Moore–Penrose generalized inverse to calculate an over-determined Mueller matrix. The process begins with the usual equation $I = AMW$, where I is the intensity matrix, A is the set of Stokes vectors from the PSG, W is the set of Stokes vectors from the PSA, and $a = \frac{1}{\sqrt{3}}$:

$$A = \begin{bmatrix} 1 & -a & a & -a \\ 1 & -a & -a & a \\ 1 & -a & -a & -a \\ 1 & -a & a & a \\ 1 & a & a & -a \\ 1 & a & -a & a \\ 1 & a & -a & -a \\ 1 & a & a & a \end{bmatrix}, \quad W = A^T, \tag{A1}$$

$$H_A = A^T A, \; H_W = WW^T, \tag{A2}$$

$$A^T I W^T = H_A M H_W, \tag{A3}$$

$$H_A^{-1} A^T I W^T H_W^{-1} = M. \tag{A4}$$

2. INVERTING A RECTANGULAR MATRIX

Here we will calculate the solution to Eq. (10):

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \tag{A5}$$

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}, \tag{A6}$$

$$\begin{bmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 6.2 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \tag{A7}$$

$$\begin{bmatrix} \frac{21}{20} \\ \frac{41}{20} \\ \frac{61}{20} \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{A8}$$

3. ΔMATRIX EXAMPLE

In this section, we will give a simple example of how the Δ matrix is calculated. To simplify the process, we will work only with a slightly over-determined 5×5 set of intensity measurements as recorded by the camera. We also have our assumed versions of A and W outlined as follows:

$$A_{as} = \begin{bmatrix} 1 & -a & a & -a \\ 1 & -a & -a & a \\ 1 & -a & -a & -a \\ 1 & -a & a & a \\ 1 & a & a & -a \end{bmatrix}, \; W_{as} = A_{as}^T. \tag{A9}$$

The intensity matrix I_m has the following values:

$$I_m = \begin{bmatrix} 1.619 & 0.398 & 0.414 & 1.375 & 1.595 \\ 0.890 & 1.264 & 1.472 & 0.523 & 0.299 \\ 1.458 & 0.750 & 1.527 & 0.455 & 0.755 \\ 0.847 & 0.928 & 0.309 & 1.543 & 1.074 \\ 1.034 & 0.521 & 0.703 & 0.793 & 1.844 \end{bmatrix}. \tag{A10}$$

From this, we can calculate M_c using $M_c = A_{as}^{-1} I_m W_{as}^{-1}$ and the Moore–Penrose inverse outlined in the previous section. This then yields the normalized Mueller matrix

$$M_c = \begin{bmatrix} 1 & 0.131 & -0.165 & -0.203 \\ 0.184 & 0.597 & -0.604 & -0.099 \\ 0.220 & 0.557 & 0.972 & 0.660 \\ 0.036 & 0.104 & -0.385 & 0.538 \end{bmatrix}. \tag{A11}$$

If we then multiply this calculated Mueller matrix by A_{as} and W_{as}, we will get I_c, i.e., $I_c = A_{as} M_c W_{as}$. The I_c matrix is outlined as

$$I_c = \begin{bmatrix} 1.574 & 0.370 & 0.457 & 1.487 & 1.630 \\ 0.841 & 1.283 & 1.533 & 0.591 & 0.292 \\ 1.484 & 0.696 & 1.663 & 0.517 & 0.796 \\ 0.931 & 0.958 & 0.328 & 1.561 & 1.125 \\ 1.052 & 0.522 & 0.741 & 0.833 & 1.903 \end{bmatrix}. \tag{A12}$$

If one compares I_c to I_m, it immediately becomes obvious that the values diverge from each other very slightly, and it is this divergence that is captured in the Δ matrix.

Funding. Horizon 2020 Framework Programme (813393).

Acknowledgment. AS acknowledges the funding for the TU/e-UU-UMCU Alliance programme.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are available in Dataset 1, Ref. [44] and Dataset 2, Ref. [45].

REFERENCES

1. R. A. Chipman, E. A. Sornsin, and J. L. Pezzaniti, "Mueller matrix imaging polarimetry: an overview," in *International Symposium on Polarization Analysis and Applications to Device Technology* (1996), pp. 5–12.
2. J. L. Pezzaniti, "Mueller matrix imaging polarimetry," Opt. Eng. **34**, 1558–1568 (1995).
3. A. De Martino, Y.-K. Kim, E. Garcia-Caurel, B. Laude, and B. Drévillon, "Optimized Mueller polarimeter with liquid crystals," Opt. Lett. **28**, 616–618 (2003).
4. F. Stabo-Eeg, "Well-conditioned multiple laser Mueller matrix ellipsometer," Opt. Eng. **47**, 073604 (2008).
5. S.-Y. Lu and R. A. Chipman, "Interpretation of Mueller matrices based on polar decomposition," J. Opt. Soc. Am. A **13**, 1106–1113 (1996).
6. M. R. Foreman and F. Goudail, "On the equivalence of optimization metrics in Stokes polarimetry," Opt. Eng. **58**, 082410 (2019).
7. D. S. Simonett, " Remote sensing with imaging radar: a review," Geoforum **1**, 61–74 (1970).
8. D. L. Evans, T. G. Farr, J. J. Van Zyl, and H. A. Zebker, "Radar polarimetry: analysis tools and applications," IEEE Trans. Geosci. Remote Sens. **26**, 774–789 (1988).
9. K. Serkowski, "Optical polarimeters in astronomy," Proc SPIE **0112**, 12–13 (1977).
10. O. Arteaga and B. Kahr, "Mueller matrix polarimetry of bianisotropic materials [Invited]," J. Opt. Soc. Am. B **36**, F72–F83 (2019).
11. E. Du, H. He, N. Zeng, M. Sun, Y. Guo, J. Wu, S. Liu, and H. Ma, "Mueller matrix polarimetry for differentiating characteristic features of cancerous tissues," J. Biomed. Opt. **19**, 076013 (2014).
12. R. M. Azzam, "Use of a light beam to probe the cell surface *in vitro*," Surf. Sci. **56**, 126–133 (1976).
13. V. A. Ushenko, B. T. Hogan, A. Dubolazov, G. Piavchenko, S. L. Kuznetsov, A. G. Ushenko, Y. O. Ushenko, M. Gorsky, A. Bykov, and I. Meglinski, "3D Mueller matrix mapping of layered distributions of depolarisation degree for analysis of prostate adenoma and carcinoma diffuse tissues," Sci. Rep. **11**, 1–12 (2021).
14. H. B. Maris, "Photo-elastic properties of transparent cubic crystals*," J. Opt. Soc. Am. **15**, 194–200 (1927).
15. F. Vachss and L. Hesselink, "Measurement of the electrogyratory and electro-optic effects in BSO and BGO," Opt. Commun. **62**, 159–165 (1987).
16. E. Slikboer, O. Guaitella, and A. Sobota, "Time-resolved electric field measurements during and after the initialization of a kHz plasma jet - from streamers to guided streamers," Plasma Sources Sci. Technol. **25**, 03LT04 (2016).
17. E. T. Slikboer, "Investigation of plasma surface interactions using Mueller polarimetry," Ph.D. thesis, Université Paris-Saclay, Eindhoven University of Technology (2018).
18. X. Tu, S. McEldowney, C. G. Yang Zou, M. Smith, N. Brock, S. S. Miller, L. Jiang, and S. Pau, "Division of focal plane red – green – blue full-Stokes imaging polarimeter," Appl. Opt. **59**, G33–G40 (2020).
19. J. Dai, F. Goudail, M. Boffety, and J. Gao, "Estimation precision of full polarimetric parameters in the presence of additive and Poisson noise," Opt. Express **26**, 34081–34093 (2018).
20. F. Goudail, "Optimal Mueller matrix estimation in the presence of additive and Poisson noise for any number of illumination and analysis states," Opt. Lett. **42**, 2153–2156 (2017).
21. J. Dai and F. Goudail, "Precision analysis of arbitrary full-Stokes polarimeters in the presence of additive and Poisson noise," J. Opt. Soc. Am. A **36**, 1229–1240 (2019).
22. B. Le Teurnier, X. Li, M. Boffety, H. Hu, and F. Goudail, "When is retardance autocalibration of microgrid-based full Stokes imagers possible and useful?" Opt. Lett. **45**, 3474–3477 (2020).
23. X. Li, H. Hu, L. Wu, and T. Liu, "Optimization of instrument matrix for Mueller matrix ellipsometry based on partial elements analysis of the Mueller matrix," Opt. Express **25**, 18872–18884 (2017).
24. Y. S. Kim, "Lorentz group in polarization optics," J. Opt. B: Quantum Semiclassical Opt. **2**, R1 (2000).
25. F. Muhammad and C. S. Brown, "Lorentz group underpinnings for the Jones and Mueller calculi," Proc. SPIE **2265**, 337–348 (1994).
26. D. Han, Y. S. Kim, and M. E. Noz, "Stokes parameters as a Minkowskian four-vector," Phys. Rev. E: Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. **56**, 6065–6076 (1997).
27. A. Peinado, A. Lizana, J. Vidal, C. Iemmi, and J. Campos, "Optimization and performance criteria of a Stokes polarimeter based on two variable retarders," Opt. Express **18**, 9815–9830 (2010).
28. A. Peinado, A. Lizana, J. Vidal, C. Iemmi, I. Moreno, J. Campos, and B. Aires, "Analysis, optimization and implementation of a variable retardance based polarimeter," in *EPJ Web of Conferences* (2010), pp. 1–7.
29. P. Lemaillet, S. Rivet, and B. Le Jeune, "Optimization of a snapshot Mueller matrix polarimeter," Opt. Lett. **33**, 144–146 (2002).
30. A. Peinado, A. Lizana, and J. Campos, "Design of polarimeters based on liquid crystals and biaxial crystals for polarization metrology," Opt. Pura Apl. **49**, 167–177 (2016).
31. R. Ossikovski, J. J. Gil, and I. San José, "Poincaré sphere mapping by Mueller matrices," J. Opt. Soc. Am. A **30**, 2291–2305 (2013).
32. I. F. Vasconcelos, R. S. De Figueiredo, S. J. De Guedes Lima, and A. S. Sombra, "Bismuth silicon oxide (Bi12SiO20-BSO) and bismuth titanium oxide (Bi12TiO20-BTO) obtained by mechanical alloying," J. Mater. Sci. Lett. **18**, 1871–1874 (1999).
33. J. S. Tyo, "Noise equalization in Stokes parameter images obtained by use of variable-retardance polarimeters," Opt. Lett. **25**, 1198–1200 (2000).
34. N. Ortega-Quijano and J. L. Arce-Diego, "Depolarizing differential Mueller matrices," Opt. Lett. **36**, 2429–2431 (2011).
35. R. Ossikovski, "Differential matrix formalism for depolarizing anisotropic media," Opt. Lett. **36**, 2330–2332 (2011).
36. R. M. Azzam, "Propagation of partially polarized light through anisotropic media with or without depolarization: a differential 4 multiplied by 4 matrix calculus," J. Opt. Soc. Am. A **68**, 1756–1767 (1978).
37. L. Vinet and A. Zhedanov, *Matrix Groups*, vol. **44** (Springer, 2011).
38. J. J. Gil and E. Bernabeu, "A depolarization criterion in Mueller matrices," Opt. Acta **32**, 259–261 (1985).
39. M. K. Swami, H. S. Patel, and P. K. Gupta, "Conversion of three by three Mueller matrix to a four by four Mueller matrix for non-depolarizing samples," Opt. Commun. **286**, 18–22 (2013).
40. O. A. Razvigor Ossikovski, "Completing an experimental nondepolarizing Mueller matrix whose column or row is missing," J. Vac. Sci. Technol. B **37**, 052905 (2019).
41. X. Li, F. Goudail, P. Qi, T. Liu, and H. Hu, "Integration time optimization and starting angle autocalibration of full Stokes imagers based on a rotating retarder," Opt. Express **29**, 9494–9512 (2021).
42. F. Goudail, X. Li, M. Boffety, S. Roussel, T. Liu, and H. Hu, "Precision of retardance autocalibration in full-Stokes division-of-focal-plane imaging polarimeters: publisher's note," Opt. Lett. **44**, 5759 (2019).
43. M. R. Foreman, A. Favaro, and A. Aiello, "Optimal frames for polarization state reconstruction," Phys. Rev. Lett. **115**, 1–6 (2015).
44. H. Philpott, E. Garcia-Caurel, O. Guaitella, and A. Sobota, "Over-determined Mueller polarimetry of overlapping pieces of BSO," figshare (2021) https://doi.org/10.6084/m9.figshare.16592744.
45. H. Philpott, E. Garcia-Caurel, O. Guaitella, and A. Sobota, "Dataset of over-determined Mueller polarimetry measurements of a quarter wave plate," figshare (2021) https://doi.org/10.6084/m9.figshare.16592753.