MRO interferometer memo: Polarisation fidelity in optical interferometers

David Buscher

Revision: 1.8 Date: 2003/08/25 22:59:59

1 Objective

This memo outlines the relationship between the polarization properties of the interferometer optics, the source polarization structure and the calibrated complex visibility signal measured by the interferometer. The worst-case errors due to source polarization and interferometer optics imperfections are estimated using simple models.

2 Summary

The product of the diattenuation of the interferometer arms and the source percentage polarization is sufficient to give an adequate measure of the leakage of signals between the Stokes Q, U, and V visibilities and the Stokes I visibility. The above statement is made proviso to the following assumptions: all interferometer beam paths are symmetric, the optics are weakly diattenuating and the sources are more strongly polarized than the calibrators.

For the articulating tertiary mirror telescope design, visibility errors due to the polarization problems are likely to be no more than about 2 percent in the most extreme imaginable case, and are less than 1 percent in more realistic cases.

3 Introduction

Oblique reflections in the optics of an interferometer mean that differently polarized light will be subject to different amplitude and phase changes when transmitted through the interferometer. Thus it is likely that sources with the same overall intensity pattern but different polarizations will give rise to different visibility measurements in the interferometer. Conversely, given a visibility measurement of a source, there will be an ambiguity between source morphology and source polarization structure, which will make scientific interpretation of the results difficult. An extra complication to this interpretation is that the polarization properties of the interferometer optics (especially the unit telescopes) will change as a function of the hour angle of the source.

One way around this is to make several measurements with different polarizers in front of the beam combiner optics and (given some knowledge of the interferometer polarization properties) use these measurements to recover both the source intensity structure and the source polarization structure. This requires complex and potentially lossy interferometer optics and severely impacts the total time required to make a given observation. A simpler alternative is to make the interferometer sensitive only to the intensity distribution and not to the polarization structure of the source. The next sections quantify what is meant by polarization structure of the source, defines a model for the polarization properties of the interferometer optics, and calculates how the two interact in an interferometric measurement.

4 Stokes images

For a single pixel detector, the Stokes parameters are defined as follows: measure the total intensity (with no polarizer present) and call this I_0. Place an ideal horizontal linear polarizer in front of the detector and call the measured intensity I_1. Rotate the linear polarizer by 45° and call the measured intensity I_2. Replace the linear polarizer with an ideal polarizer which lets though only left-handed circular polarization and call the measured intensity I_3. Then the Stokes parameters are

$$S_0 = I_0 \quad (1)$$
$$S_1 = I_1 - I_0 \quad (2)$$
$$S_2 = I_2 - I_0 \quad (3)$$
$$S_3 = I_3 - I_0 \quad (4)$$

These parameters are sometimes labeled as I, Q, U and V, but we adopt the above convention here (V gets confused with visibility) — the main thing to remember is that "Stokes I" corresponds to S_0.

The single-pixel Stokes parameters can readily be generalized to a spatially-varying intensity distribution (which we shall call here an "image") where $I_0(x, y)$, $I_1(x, y)$ etc represent the intensities measured at the pixel with coordinate (x, y). Any incoherent partially polarized intensity distribution is then *fully determined* by the set of four images $\{S_0(x, y), S_1(x, y), S_2(x, y), S_3(x, y)\}$. An ideal polarimetric interferometer will measure all four of these images. Alternatively, a "polarization fidelity" interferometer will measure only $S_0(x, y)$, but measure this independent of the values of $S_1(x, y)$, $S_1(x, y)$ and $S_3(x, y)$: in other words a polarization fidelity interferometer will minimize the "leakage" of the S_1, S_2 and S_3 images into the measurement of the S_0 image.

5 Interferometer optics

For simplicity, we analyze here an optical interferometer which is not affected by atmospheric turbulence, i.e. where the complex fringe visibility is a good observable. It is relatively easy to use the results presented here to derive results which include the effects of atmospheric turbulence, since the atmosphere is non-polarizing.

A beam entering one telescope, traveling through the interferometer optics, and landing on the fringe detector undergoes a polarization state transformation as it passes through or reflects off each optical surface. If we assume for simplicity a quasi-monochromatic plane wavefront and spatially homogeneous optical surfaces, then the instantaneous polarization state of the beam can be described by a simple Jones vector $J = [E_x, E_y]$ and the polarization transformation at any surface can be described by a Jones matrix M_i for surface i which acts on the input Jones vector to produce an output Jones vector. It follows that there exists a Jones matrix for any given arm of the interferometer $M = \prod_{i=1}^{n} M_i$ which describes the transformation of the light through the entire optical train to the detector.

For any optical system with Jones matrix M there will be a corresponding pair of orthonormal Jones vectors J_+ and J_- which represent *characteristic polarization states*. These states are eigenstates of the optical system, i.e they pass through the optics without a change in the state of polarization:

$$MJ_+ = a_+ J_+ \quad (5)$$
$$MJ_- = a_- J_- \quad (6)$$

where a_+ and a_- are scalar complex eigenvalues with $|a_+|, |a_-| < 1$ for a lossy optical system. Note that these eigenstates are in general elliptical polarization states: linear or circular polarization states are special cases.

We hereafter assume that the polarization properties of the different interferometer arms are identical, which can be achieved by making all the beam paths symmetric with respect to number of surfaces and their incidence angles (this is not possible in the beam combiner optics, but the effects of any asymmetries can be minimized by designing the beam combiner to have only low angles of incidence). As a result the Jones matrices of the two arms will be identical (we can multiply the Jones matrix for each arm by a different complex scalar to encode any optical path delay differences between arms, but choose not to do so here for simplicity).

In such a case, the interference of two partially coherent beams can be fully characterized by the interference of each of the two characteristic polarizations from one arm of the interferometer with the corresponding polarization state from the other arm of the interferometer: the two characteristic polarizations are orthogonal so there is no cross-interference term. If the instantaneous E-fields incident on each of two telescopes in each of the two polarizations are denoted by $E_+(1)$, $E_-(1)$, $E_+(2)$ and $E_-(2)$ respectively then the complex amplitude of the fringes at the beam combiner will be given by

$$A_{12} = a_+(1)a_+^*(2) < E_+(1)E_+^*(2) > + a_-(1)a_-^*(2) < E_-(1)E_-^*(2) >$$

Noting that the two arms are identical in terms of polarization so that $a_+(1) = a_+(2) = a_+$ we have

$$A_{12} = |a_+|^2 < E_+(1)E_+^*(2) > + |a_-|^2 < E_-(1)E_-^*(2) > \ . \quad (7)$$

This equation can be interpreted as saying that the observed fringe pattern is the superposition of two intensity patterns corresponding to the interference of the E_+ fields from the two interferometer arms and the interference of the E_- fields from the two arms respectively. We note the important result that *with a symmetric optical system only the modulus of the polarization transfer coefficients a_+ and a_- have any effect on the interference pattern — the phase shifts between polarizations have no effect.*

6 Interferometric visibility measurement

A general interferometric beam combiner and its associated detectors will output a set of discrete intensity measurements $\{I_j : j = 1 \ldots n\}$ with a characteristic modulation in space or time which encodes the fringe visibility. We assume that the complex fringe amplitude on a particular baseline can be demodulated from the intensity measurements through a weighted sum (e.g. using a discrete Fourier transform), expressed as

$$A(u_i, v_i) = \sum_{j=1}^{n} w_{ij} I_j \ , \quad (8)$$

where (u_i, v_i) is the angular frequency of the sinusoidal component of the sky brightness distribution which is measured by interferometer baseline i and $\{w_{ij} : j = 1 \dots n\}$ is a set of complex weight factors. Denoting the total flux as

$$A(0) = \sum_{j=1}^{n} w_{0j} I_j \tag{9}$$

(where in most cases $w_{0j} = 1$), we define the (uncalibrated) complex fringe visibility as

$$V(u_i, v_i) = A(u_i, v_i)/A(0) \tag{10}$$

Now, each of the intensity measurements I_j can be expressed as

$$I_j = I_{j,+} + I_{j,-} \tag{11}$$

where $I_{j,+}$ and $I_{j,-}$ are the intensities which would be measured when a pair of perfect polarizers, chosen to select the characteristic polarizations of the interferometer arms, are placed in turn in front of the fringe detector. By combining equations 8 and 11, we can express the measured fringe amplitude as the sum of the fringe amplitudes in the two polarizations:

$$A(u_i, v_i) = A_+(u_i, v_i) + A_-(u_i, v_i) \tag{12}$$
$$A(0) = A_+(0) + A_-(0) \tag{13}$$

where the notation should be self-explanatory.

7 Evaluating the polarization fidelity of an interferometer

A polarization fidelity interferometer is one which gives the same complex visibility measurement for a given $S_0(x, y)$ independent of the values of $S_1(x, y)$, $S_2(x, y)$ and $S_3(x, y)$. One obvious example of a such an interferometer is an interferometer in which the state of polarization of the beams is unchanged after passing through the interferometer optics. In such a case it is clear that the interferometer setup corresponds to measuring an intensity pattern in Stokes S_0. Any interferometer which gives the same visibility measurement as such a perfect interferometer is clearly a polarization fidelity interferometer. In this section we use evaluate the polarization fidelity of an arbitrary interferometer by comparing it with the visibility measured on the same source using a perfect interferometer.

In the case of a perfect interferometer it is easy to show that *any* two orthonormal Jones vectors J_+ and J_- represent characteristic states of the interferometer optics. We can choose these states to be the same two states which are the characteristic states of a given imperfect interferometer (at a given moment in time — these states will in general change as the hour angle of the source being observed changes). The only difference between the perfect interferometer and the imperfect interferometer is then the fact that for a perfect interferometer the eigenvalues a_+ and a_- are both exactly unity, whereas in an imperfect interferometer they are not.

It can be shown that we can analyze the light from an arbitrary partially polarized source in terms of the second-order statistics of the instantaneous electric field measured in any two orthogonal polarization states. Thus it is helpful to split the object being observed into two images, denoted as $S_+(x, y)$ and $S_-(x, y)$ which correspond to the images which would be seen through polarizers selecting the J_+ and J_- states respectively. In general these are some linear combinations of the S_0, S_1, S_2 and S_3 images. We can consider the light in each of the two characteristic states passing separately through the interferometer optics and we see that the corresponding fringe amplitudes observed in the two states are given by

$$A_+(u_i, v_i) = |a_+|^2 \tilde{S}_+(u_i, v_i) \tag{14}$$
$$A_-(u_i, v_i) = |a_-|^2 \tilde{S}_-(u_i, v_i) \tag{15}$$

where $\tilde{S}(u_i, v_i)$ corresponds to the Fourier component of $S(x, y)$ on a baseline (u_i, v_i). Combining equations 12, 14 and 15 we have

$$A(u_i, v_i) = |a_+|^2 \tilde{S}_+(u_i, v_i) + |a_-|^2 \tilde{S}_-(u_i, v_i) \tag{16}$$

and we can similarly derive that

$$A(0) = |a_+|^2 \tilde{S}_+(0) + |a_-|^2 \tilde{S}_-(0) \tag{17}$$

The measured fringe visibility is therefore given by

$$V(u_i, v_i) = \frac{A(u_i, v_i)}{A(0)} = \frac{|a_+|^2 \tilde{S}_+(u_i, v_i) + |a_-|^2 \tilde{S}_-(u_i, v_i)}{|a_+|^2 \tilde{S}_+(0) + |a_-|^2 \tilde{S}_-(0)} \tag{18}$$

In order to quantify the polarization fidelity of a given interferometer we express the measured visibility relative to that which would be measured for the same source by a perfect interferometer $V_0(u_i, v_i)$:

$$\frac{V(u_i, v_i)}{V_0(u_i, v_i)} == \frac{|a_+|^2 \tilde{S}_+(u_i, v_i) + |a_-|^2 \tilde{S}_-(u_i, v_i)}{|a_+|^2 \tilde{S}_+(0) + |a_-|^2 \tilde{S}_-(0)} \times \frac{\tilde{S}_+(0) + \tilde{S}_-(0)}{\tilde{S}_+(u_i, v_i) + \tilde{S}_-(u_i, v_i)} . \tag{19}$$

We define the diattenuation D of an optical system as the fractional difference in transmission of the two states

$$D = \frac{|a_+|^2 - |a_-|^2}{|a_+|^2 + |a_-|^2} , \tag{20}$$

where we have adopted the convention that $|a_+| \geq |a_-|$. A little algebraic manipulation yields the visibility relative to a perfect interferometer in terms of the diattenuation:

$$\frac{V(u_i, v_i)}{V_0(u_i, v_i)} = \frac{1 + D \left\{ \frac{\tilde{S}_+(u_i,v_i) - \tilde{S}_-(u_i,v_i)}{\tilde{S}_+(u_i,v_i) + \tilde{S}_-(u_i,v_i)} \right\}}{1 + D \left\{ \frac{\tilde{S}_+(0) - \tilde{S}_-(0)}{\tilde{S}_+(0) + \tilde{S}_-(0)} \right\}} . \tag{21}$$

This equation is the main mathematical result of this memo. We identify the term $\left\{ \tilde{S}_+(u_i, v_i) - \tilde{S}_-(u_i, v_i) \right\}$ as a Fourier component of the polarization difference image $\{S_+(x, y) - S_-(x, y)\}$, while $\left\{ \tilde{S}_+(u_i, v_i) + \tilde{S}_-(u_i, v_i) \right\}$ is a Fourier component at the same frequency of the unpolarized flux $S_0(x, y)$. Thus the ratio of the two can be thought of as a "Fourier percentage polarization". The corresponding term in the denominator which is the same expression evaluated for the "zero spacing", i.e. the total flux.

8 Practical examples

For the particular case of an unpolarized source we have $S_+(x, y) = S_-(x, y) = \frac{1}{2} S_0(x, y)$ and so a visibility measurement gives $V(u_i, v_i) = V_0(u_i, v_i)$ which leads to the important result that *for* *an unpolarized source, we measure the same visibility independent of the interferometer optics polarization properties*. This is of particular relevance for calibration of visibility measurements: most calibrators will be normal stars and will not be very resolved, so the object polarization will be low (typically $<<$ 1 percent). As a result, when deriving the effects of interferometer polarization on a *calibrated* visibility measurement, we only need to consider the polarization properties of the optics when observing the source (which potentially is significantly polarized), and not those when observing the calibrator.

In general the Jones matrix M of an interferometer will be rather complex to analyze, but we can restrict analysis to a few simple cases, with the hope that these are indicative of the general behavior. The most obvious simplification is to only analyze interferometer geometries for which the linear polarization states are the eigenstates. This is particularly easy to interpret in terms of the source polarization, since for most astronomical observations at optical wavelengths, processes giving rise to linear polarization are dominant over processes giving rise to circular polarization.

If the beam relay geometry is horizontal and planar, then the beam relay eigenstates are the horizontal and vertical linear polarization states. If the unit telescopes are pointing in directions such that the S and P directions for any oblique reflections within the telescope are the same as those for the rest of the beam train, then the horizontal and vertical polarizations are eigenstates for the entire system. For the articulating tertiary telescope design (shown in figure 1) this will occur for any orientation of the inner rotation axis (i.e. the "pitch axis"), provided that the outer gimbal (i.e. the one which rotates about the "roll axis") is horizontal.

For the interferometer beam relay and combination optics the angle of incidence of the starlight beam on the optics is less than or equal to 30 degrees, and for such angles the diattenuation for silver coated surfaces is small ($<$ 0.5 percent), so we consider here only the oblique reflections in the unit telescopes, which potentially have larger angles of incidence. Calculations by Chris Haniff show that, for the articulating tertiary with a silver-coated tertiary mirror, the diattenuation is less than 1 percent when observing sources above 30 degrees elevation for all wavelengths longer than about 800nm. For observations at 600nm and extreme angles, the diattenuation rises to slightly over 2 percent, so we will use 2 percent as a worst-case diattenuation for illustration purposes.

The percentage linear polarization of the source enters into both the numerator and the denominator of equation 21. The $[S_+(0) - S_-(0)] \, / \, [S_+(0) + S_-(0)]$ term in the denominator refers to the "zero spacing" polarized flux i.e. the polarization integrated over the seeing disk. For almost all astronomical sources, this polarization at most a few percent at visible wavelengths, and so the denominator is unity to better than a few parts in 10^4; we take it to be unity for the rest of this discussion.

The $[S_+(u_i, v_i) - S_-(u_i, v_i)] \, / \, [S_+(u_i, v_i) + S_-(u_i, v_i)]$ term in the numerator, which depends on the polarization at high spatial frequencies, is less easy to estimate *a priori* since such properties have not been measured extensively. Sources are expected to be more polarized on small angular scales than they are on larger scales, but little theoretical work has been done on the expected polarization structure on milliarcsecond scales. We shall use here two "toy" astronomical models and examine the worst-case effects on the polarization fidelity.

The first model is a simple starspot model, consisting of a uniform unpolarized disk and an unresolved spot which emits 10 percent of the flux of the star and is 100 percent linearly polarized in a direction which happens to match the J_+ characteristic polarization of the interferometer. On short baselines which do not resolve the disk, $[S_+(u_i, v_i) - S_-(u_i, v_i)] \, / \, [S_+(u_i, v_i) + S_-(u_i, v_i)]$ is of order 10 percent, so the fractional error in the visibility is typically of order 0.2 percent. On long baselines where the disk is fully resolved, the visibility of the spot dominates, and

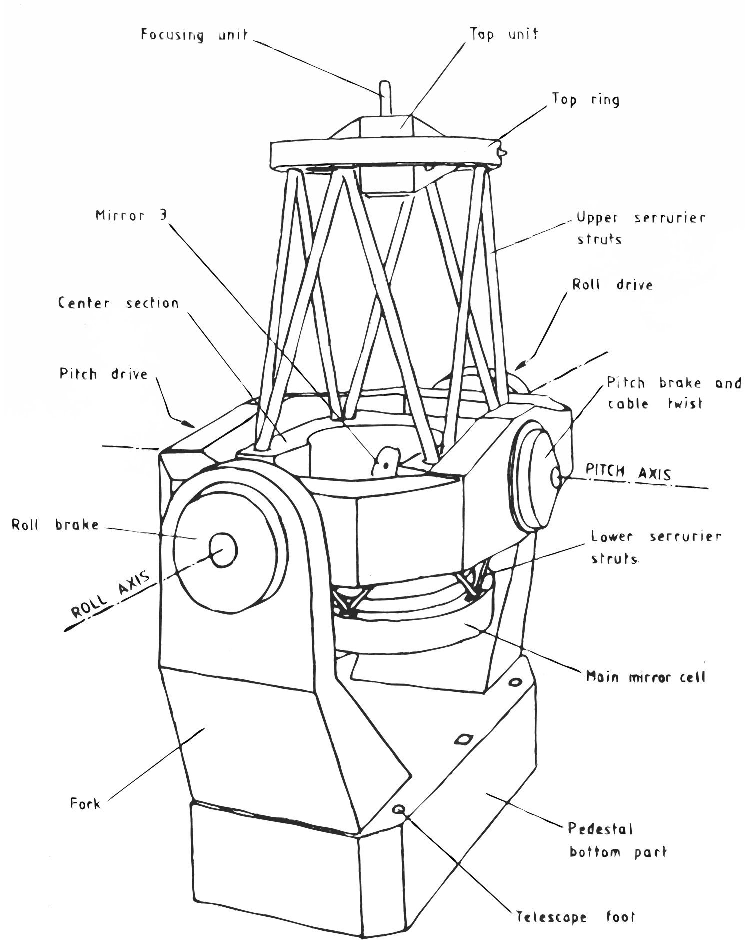

Figure 1: Layout of an articulating-tertiary telescope showing the orientation of the pitch and roll axes. For an interferometer unit telescope, the collimated beam exits through the roll axis. The diagram shows the CAT telescope (diagram courtesy of ESO).

so $S_-(u_i, v_i) \simeq 0$. This yields *fractional* errors in the measured visibility of order 2 percent. However, the fringe visibility on these baselines is of order 10 percent, so the actual visibility errors are of order 0.2 percent.

A second, more extreme, model consists of a binary star in which one star is 100 percent linearly polarized in the J_+ direction and the other star is equally bright and 100 percent linearly polarized in the J_- direction. In such a case, the fringe contrast measured by a perfect polarization-fidelity interferometer will go to zero when the projected baseline is such that $S_+(u_i, v_i) = -S_-(u_i, v_i)$. On this baseline, the fractional visibility error of *any* imperfect interferometer is infinite. However, any error due to polarization leakage must be compared to other potential sources of error. Error sources such as photon noise give rise to measurement errors which are finite when the fringe visibility is zero and so ultimately we need to consider the absolute as well as fractional visibility errors.

For the example being considered here, the absolute visibility can be computed by evaluating

$$V(u_i, v_i) = \left(\frac{\tilde{S}_+(u_i, v_i) + \tilde{S}_-(u_i, v_i)}{\tilde{S}_+(0) + \tilde{S}_-(0)}\right)\left(\frac{1 + D\left\{\frac{\tilde{S}_+(u_i,v_i) - \tilde{S}_-(u_i,v_i)}{\tilde{S}_+(u_i,v_i) + \tilde{S}_-(u_i,v_i)}\right\}}{1 + D\left\{\frac{\tilde{S}_+(0) - \tilde{S}_-(0)}{\tilde{S}_+(0) + \tilde{S}_-(0)}\right\}}\right) \quad (22)$$

$$= V_0(u_i, v_i) + \frac{D\left\{\tilde{S}_+(u_i, v_i) - \tilde{S}_-(u_i, v_i)\right\}}{\tilde{S}_+(0) + \tilde{S}_-(0)} \quad (23)$$

where we have made use of the fact that $\tilde{S}_+(0) = \tilde{S}_-(0)$ (although this condition holds approximately for most cases of astronomical interest, it holds exactly here). In the case of the binary star, evaluation of equation 23 shows that the absolute visibility error is 2 percent. Astronomical sources with this polarization structure are rather unlikely to occur in nature so this can be seen to be a very pessimistic upper bound.