ASTRONOMY & ASTROPHYSICS
SUPPLEMENT SERIES

Astron. Astrophys. Suppl. Ser. **117,** 137-147 (1996)

Understanding radio polarimetry. I. Mathematical foundations

J.P. Hamaker[1], J.D. Bregman[1] and R.J. Sault[1,2]

[1] Netherlands Foundation for Research in Astronomy, Postbus 2, 7990 AA Dwingeloo, The Netherlands
[2] Australia Telescope National Facility, CSIRO, P.O. Box 76, Epping, N.S.W. 2121, Australia

Received August 4; accepted October 22, 1995

Abstract. — The measurement of polarized radiation uses entirely different methods at optical and radio wavelengths. As a result, the algebraic analysis of polarimeter performance differs and, in the case of radio interferometry, is unnecessarily complicated. We demonstrate that the mathematical operation of outer matrix multiplication provides the missing link between the two approaches. Within one coherent framework, we then unite the concepts of Stokes parameters and Wolf coherency matrix, the Jones and Mueller calculi from optics, and the techniques of radio interferometry based on multiplying correlators. We relate the polarization performance of a complete radio interferometer to the (matrix) polarization properties of its successive signal processing stages, providing a clear view of how a radio polarimeter works. Our treatment also clarifies the nature of and the relations between the various types of transformations used in optical polarimetry. We develop the analysis from the radio interferometrist's point of view, but include enough background for a wider audience. In a companion paper, we discuss in more detail the application to the calibration of radio interferometer systems; in a third paper we investigate the IAU (1973) radio definition of the Stokes parameters and its precise translation into mathematical form.

Key words: methods: analytical — methods: data analysis — techniques: interferometers — techniques: polarimeters

1. Introduction

The study of polarized electromagnetic radiation belongs traditionally to the realm of optics. A century and a half ago, the representation of polarization in terms of what are now known as the four Stokes parameters was discovered (Stokes 1852; modern treatments are to be found in most advanced optics and radio-astronomy textbooks, such as Born & Wolf 1964; Hecht & Zajac 1982; Thompson et al. 1986). In modern optical-instrument design, these parameters are treated as the components of the Stokes vector, whose transmission through an optical element may be described by multiplication with a 4×4 *Mueller matrix* (Mueller 1948). Another approach is to consider the transmission of the instantaneous (complex) optical vector amplitude; this formulation describes an optical element by multiplication with a 2×2 *Jones matrix* (Jones 1941). The Jones formalism assumes the quasi-monochromatic case (cf. Sect. 6) and is, in its simplest form, less generally applicable than the Mueller formalism. Several authors indicate a connection between the Jones and Mueller matrices which seems to be due originally to van de Hulst (1957).

Unlike optical detectors, radio antennas are inherently fully polarized. Observers were therefore forced to consider polarization from the very beginning. The instrumental techniques at radio wavelengths are too different for the theories from optics to be directly applicable, so radio observers had to develop a theory of their own. The foundation was laid by Morris et al. (1964). They consider two antennas, each with a single *receptor*, connected to the two inputs of a device that *correlates* the input signals, i.e. multiplies them and time-averages the product. They derive a lengthy formula showing the output in terms of the *Stokes visibilities* (Appendix C) with the *orientations* and *ellipticities* of the dipoles (Born & Wolf 1964; Thompson et al. 1986) as parameters in an otherwize error-free system. We shall call this the *black-box formula* because it treats the entire interferometer as a black box that converts input Stokes visibilities into an output correlation coefficient. All observational work since the appearance of this formula has been based, directly or indirectly, upon it. Later theoretical work uses it as a starting point for considering the impact of various instrumental errors (e.g. Weiler 1973; Weiler & Raimond 1976; Thompson et al. 1986).

Send offprint requests to: J.P. Hamaker, jph@nfra.nl

As a result of these efforts, radio polarimetrists now understand quite well what they are doing. Yet a comprehensive framework in which all elements of the theory find their natural place is lacking: The basic problem remains that the black-box formula describes only the interferometer in its entirety without providing an insight into its inner workings.

This paper has several purposes. First and foremost, it presents a theory that unifies the concepts of Stokes parameters, Jones and Mueller matrices and of radio interferometry in a single coherent framework. The key to this unification is the observation (made earlier by opticists but never brought to bear on radio polarimetry) that the *coherency matrix* (Born & Wolf 1964) can be reformatted as a coherency *vector* that is the *outer matrix product* of two vector amplitudes á la Jones. The elegant properties of the outer product then lead the way to a simple and elegant theory.

In developing this theory, we emphasize some basic concepts from the theory of linear algebra which are well known in theory but appear to be often overlooked in practice. Our second aim, then, is to present the theory in a semi-tutorial style that emphasizes the precise interpretation of the mathematics. We start from the conceptual framework of radio interferometry, but we try to explain enough of the basics from that field to make the material accessible to non-specialists.

In the third place, we indicate how previous work on radio interferometry fits into the new framework.

Our theory is based on the theory of *linear* transformations. Like practically all previous work on radio polarimetry, it excludes one important aspect of practical radio polarimetry, viz. the variation of instrumental polarization over the primary beam of the element antennas which is induced by the feed/reflector geometry.

In a companion paper (Paper II, Sault et al. 1996) of a more specialist nature, we consider the application of our theory to calibration problems of practical radio interferometer arrays. That paper relies both on results from the optical-polarimetry literature and on the material in this paper. One important difference may be worth noting from the outset: Here we treat the processing of the signal and coherency vectors in a single two-element interferometer. Paper II provides a broader view by investigating the properties of a multi-element interferometer array as an imaging device.

In a third paper (Paper III, Hamaker & Bregman 1996) we consider the interpretation of the IAU (1973) radio definitions of the Stokes parameters for radio astronomy and seek to definitively clear up the confusion that seems to beset their application. In the conventions it adopts, the present paper anticipates the results of Paper III, which are also in accord with the conventions used by Thompson et al. (1986).

2. Interferometer block diagram

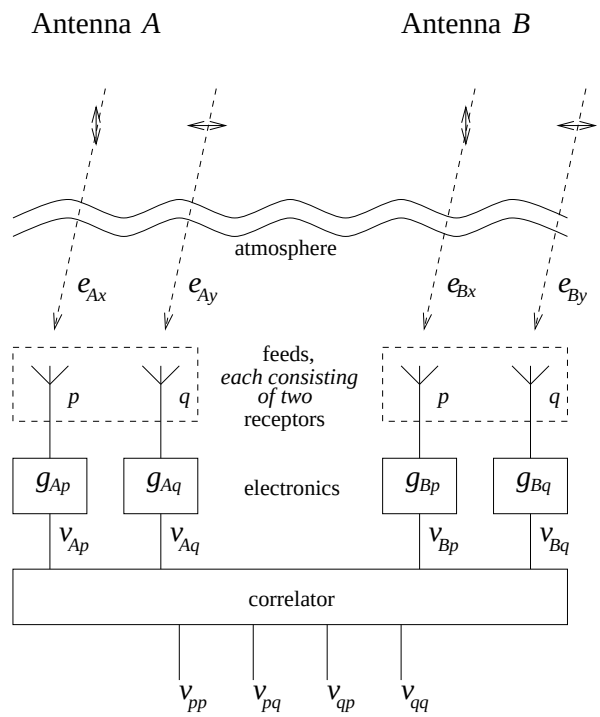

Fig. 1. Interferometer block diagram

Figure 1 gives a schematic diagram of a polarization interferometer, demonstrating the terminology we will use. The incident electric field is described in terms of two components (e.g. horizontal and vertical as shown in the diagram, or right- and left-circular), each of which is converted to an electric voltage by a *receptor*. In a typical *feed*, the centres of the two receptors coincide geometrically.[1] The receptors are usually labeled according to the polarization component to which they are sensitive, i.e. X and Y for linearly polarized receptors, R and L for circularly polarized ones; to avoid such specific connotations, we label them p and q. The interferometer consists of two *antennas* which we label A and B, connected to a *correlator* that measures the four cross-correlations between a voltage from antenna A and one from antenna B (cf. Eq. (10) below); this correlator is the full-polarization equivalent of the multiplier mentioned in Sect. 1. The complete signal chain from the atmosphere down to the correlator inputs is called an *interferometer arm*.

We note in passing that this model includes the single-dish polarimeter as a special case, viz. where the two telescopes A and B coalesce into one.

[1] Mechanical constraints sometimes force them some small distance apart along the line of sight; in the paraxial approximation used here such a shift is mathematically described by a phase factor.

3. The mathematical model of an interferometer

3.1. *The signal in a single interferometer arm*

A *quasi-monochromatic* signal propagating in space may be characterized at some fixed point by its *instantaneous complex vector amplitude* $\boldsymbol{e}(t)$. Along its path, the vector undergoes transformations, each of which we represent by an *operator* $\boldsymbol{J}$:

$$\boldsymbol{e}_{\rm out}(t) = \boldsymbol{J}\boldsymbol{e}_{\rm in}(t) \tag{1}$$

The limits of validity of this representation are discussed in Sect. 6. We assume all operators of interest to be *linear*; we also henceforth drop the explicit dependence on time t and refer to $\boldsymbol{e}$ as the *signal vector*.

It is common practice to represent the vectors $\boldsymbol{e}$ as columns of numbers, the *coordinates*; in this representation, Eq. (1) becomes a matrix multiplication. Note, however, that such a representation implicitly assumes a *coordinate system* in which the coordinates are measured. The distinction between a vector *per se* and its many possible representations becomes important when we want to deal with more than one coordinate system, - as is the case here.

In an xyz coordinate system with the z axis along the direction of propagation

$$\boldsymbol{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix}$$

The matrices $\boldsymbol{J}$ describe the optical elements in the path. They are known in optics as *Jones matrices* after their inventor (Jones 1941; Azzam & Bashara 1987).

We now generalize the concept of a geometrical signal path to include the feed and electronics of the antenna. We *define* an *electronics coordinate system* by postulating that the voltages $v_{\rm p}, v_{\rm q}$ are the Cartesian components of an electronic signal vector $\boldsymbol{v}$. The transition from electromagnetic radiation to electric voltage pair occurs at the feed:

$$\boldsymbol{v} = \boldsymbol{Q}\boldsymbol{e} \tag{2}$$

In terms of physical dimensions, $\boldsymbol{Q}$ converts an electric field (V/m) to a voltage (V), so it has the dimension of a length.

3.2. *The coherency vector*

The coherency properties of the electric field are described by the *coherency matrix* (Born & Wolf 1964). An equivalent form that is more convenient for our purposes is the coherency *vector*; in xy coordinates it is defined as

$$\boldsymbol{e} =< \begin{pmatrix} e_{{\rm A}x}e^*_{{\rm B}x} \\ e_{{\rm A}x}e^*_{{\rm B}y} \\ e_{{\rm A}y}e^*_{{\rm B}x} \\ e_{{\rm A}y}e^*_{{\rm B}y} \end{pmatrix} > \tag{3}$$

This vector form enables us to make the crucial step of recognizing the coherency vector as the time-averaged *outer product* of the two input signal vectors:

$$\boldsymbol{e} =< \boldsymbol{e}_{\rm A} \otimes \boldsymbol{e}^*_B > \tag{4}$$

Note that in Eq. (4) we are introducing a new notational convention to distinguish quantities in the 2-dimensional *signal domain* from those in the 4-dimensional *coherency domain*: The former refer to a single antenna and must therefore be subscripted with either A or B; the latter refer to the interferometer as a whole and need no such subscripts.

The outer product is also known under other names, e.g. *direct, tensor* or *Kronecker* product. A brief description is given in Appendix A. Its most important property for our purposes is that for any four matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{A}', \boldsymbol{B}'$ of appropriate dimensions the following identity holds:

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{A}' \otimes \boldsymbol{B}') \;=\; (\boldsymbol{A}\boldsymbol{A}') \otimes (\boldsymbol{B}\boldsymbol{B}') \tag{5}$$

The outer product provides us the missing link between the transformations of signals in the individual interferometer arms and those of the coherency vector in the interferometer as a whole. Indeed, consider some transforming element $\boldsymbol{J}$ in each of the individual antenna paths:

$$\boldsymbol{e}_{\rm A,out} = \boldsymbol{J}_{\rm A}\boldsymbol{e}_{\rm A,in}\ ; \quad \boldsymbol{e}_{\rm B,out} = \boldsymbol{J}_{\rm B}\boldsymbol{e}_{\rm B,in}$$

Then

$$\begin{aligned} \boldsymbol{e}_{\rm out} &=< \boldsymbol{e}_{\rm A,out} \otimes \boldsymbol{e}^*_{\rm B,out} > \\ &= (\boldsymbol{J}_{\rm A} \otimes \boldsymbol{J}^*_{\rm B}) \ < (\boldsymbol{e}_{\rm A,in} \otimes \boldsymbol{e}^*_{\rm B,in}) > \\ &= \boldsymbol{J}\ \boldsymbol{e}_{\rm in} \end{aligned} \tag{6}$$

For simplicity, we henceforth drop the <> averaging signs. The assumption that we make in doing so will be investigated in Sect. 6.

We emphasize that the coherency vector, too, exists *independently* of the specific coordinate system that we used to define it (cf. Appendix A). The Cartesian form Eq. (3) is just a specific representation; we call it the *geometric xy* representation.

3.3. *Coordinate transformations*

So far, we have described our vectors in terms of a specific geometric coordinate system defined by x, y and z axes. If we had chosen a different system, the same vector would be *represented* by a different array of values. Given the representation in one Cartesian system, one finds that in another one through a matrix multiplication of the same form as Eq. (1):

$$\boldsymbol{x}_{\rm new} = \boldsymbol{T}\boldsymbol{x}_{\rm old}$$

Along with the representation of all vectors, $\boldsymbol{T}$ also transforms the representation of linear operators such as $\boldsymbol{J}$ in Eq. (1):

$$\boldsymbol{J}_{\rm new} = \boldsymbol{T}\boldsymbol{J}_{\rm old}\boldsymbol{T}^{-1} \tag{7}$$

In Sect. 3.1, Eq. (1) represents a physical transformation of the signal vector. The same expression may also represent a linear *coordinate* transformation. We will encounter both types of transformations in practice. Where necessary, we will use superscripts to indicate in which particular coordinate system a vector or operator is represented: a '+' for geometric xy coordinates, a '$\odot$' for circular-polarization coordinates and an S for Stokes coordinates (see Sect. 3.4 below).

Coordinate transformations between Cartesian systems are represented by *unitary* matrices, which have the useful property that they can be inverted by transposition and complex conjugation:

$$\boldsymbol{X}^{-1} = \boldsymbol{X}^{*T}$$

The Cartesian coordinate systems of interest are not limited to those representing physical space. Abstract coordinate frames play a prominent role in polarization theory, e.g. the circular-polarization frame in the signal domain and the Stokes frame in the coherency domain.

As a final point, we observe that formally there is no difference between the two types of transformation. In some cases this is physically obvious, e.g. a Faraday rotation over an angle χ is equivalent to rotating the feed over $-\chi$. In other cases there is no such obvious equivalence.

One may ask whether the difference between the two types of transformation is at all relevant if they are equivalent in their mathematical expression. We suggest that it is, because it provides an important clarifying perspective on the physics of the problem that is lacking in most existing work.

3.4. *The Stokes representation of the coherency vector*

It is customary to analyse the source in terms of *Stokes visibilities* I, Q, U, V, which in combination form a vector, the *Stokes vector*; the connection with the more familiar Stokes parameters from optics is outlined in Appendix C. In terms of the xy representation of the coherency vector, $\boldsymbol{e}^+$, the Stokes vector is defined by

$$\boldsymbol{e}^{\mathrm{S}} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \boldsymbol{T}\boldsymbol{e}^+; \quad \boldsymbol{T} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -i & i & 0 \end{pmatrix} \tag{8}$$

From the viewpoint developed above, the Stokes vector is not really a vector in its own right. It is just the representation of the coherency vector in an abstract frame, that of *Stokes coordinates*, linked to the geometric xy frame through the coordinate transformation $\boldsymbol{T}$. The notation $\boldsymbol{e}^{\mathrm{S}}$ introduced in Eq. (8) for the Stokes vector emphasizes this view.

$\boldsymbol{T}$ is unitary except for a missing normalizing factor. In this paper and its companions we will only use the inverse of $\boldsymbol{T}$, $\boldsymbol{S} = \boldsymbol{T}^{-1}$, which we show here for future reference:

$$\boldsymbol{S} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & -i \\ 1 & -1 & 0 & 0 \end{pmatrix} \tag{9}$$

4. The signal path in an interferometer

A radio interferometer measures the coherency by literally perfoming the correlation operations of Eq. (3): It converts the radiation vector amplitudes at two antennas A and B into voltage amplitudes and submits these to a *correlator* that multiplies and averages these voltages to deliver

$$\boldsymbol{v} = < \begin{pmatrix} v_{\mathrm{A}x} v^*_{\mathrm{B}x} \\ v_{\mathrm{A}x} v^*_{\mathrm{B}y} \\ v_{\mathrm{A}y} v^*_{\mathrm{B}x} \\ v_{\mathrm{A}y} v^*_{\mathrm{B}y} \end{pmatrix} > \tag{10}$$

In the path between the source and the correlator inputs, a number of intentional and spurious signal transformations occur. The remainder of this paper is concerned with the description of these transformations and their effects.

In a practical correlator, the averaging occurs over a number of variables including some that are outside the scope of this paper. We will consider this matter a little further in Sect. 6.

4.1. *The interferometer equation*

We consider a more detailed model of one interferometer arm (antenna A) with the various elements that transform the signal. A schematic is given in Fig. 2. The incident electromagnetic signal is subject to the following transformations:

- Faraday rotation in the Earth's ionosphere, $\boldsymbol{F}$.
- A possible rotation $\boldsymbol{P}$ over a *parallactic angle* of the feed system with respect to the incident field. This is a transformation from the coordinate system of the rotating sky to the system of the rotating antenna mount on which the feed rides. For an equatorial mount the parallactic angle is zero.
- If one were to consider the *primary-beam polarization* caused by diffraction on the feed/reflector geometry of our antennas, the appropriate transformation would be placed here. The transformation would be a different one for each different source position in the primary beam, so for a distributed source we cannot represent it in the simple form of Eq. (1). This is a fundamental problem in radio interferometery that is only beginning to be addressed by some observatories; here we simply declare it to be outside the scope of this paper.
  Another effect in the reflector geometry that we omit from our equations is the sign reversal of Stokes V that occurs in each reflection (Simmons & Guttman 1970).

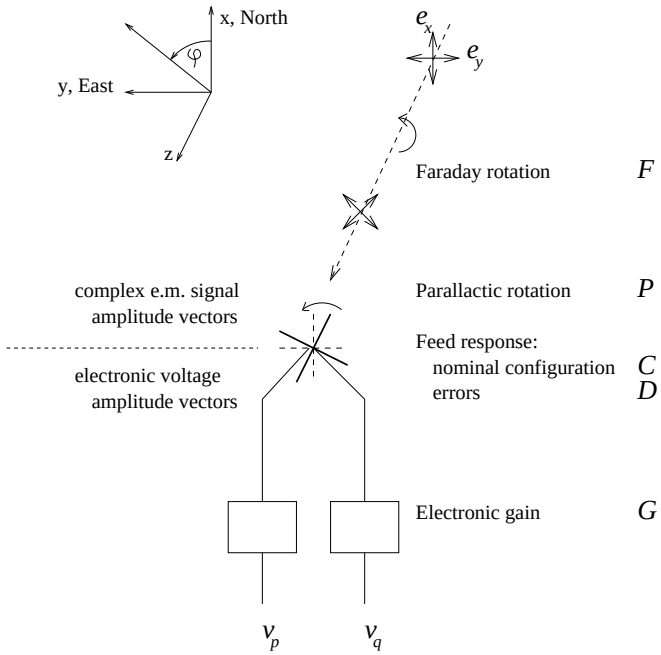

Fig. 2. Schematic of the coordinate and vector transformations. The orientations of the x and y axes and the direction of positive rotation shown are in accordance with the IAU radio definition (cf. Paper III)

- The feed, $\boldsymbol{Q}$ of Eq. (2), may be quite complex to model in detail. One may reduce it, however, to a product of two matrices:
  - An idealised *nominal* feed configuration, $\boldsymbol{C}$: A coordinate transformation from the frame of the rotating antenna mount to the electronic-voltage frame.
  - The deviations of the actual feed from the ideal, $\boldsymbol{D}$. For an error-free feed, $\boldsymbol{D} = \boldsymbol{I}$, the identity matrix.
  - Some systems include a *hybrid* that converts the outputs from a linearly polarized feed into circular components. The combination $\boldsymbol{HDC}$ effectively forms a circular feed with its errors. Appendix **D**.1 shows how we can convert this form to the combination of a configuration matrix and a feed matrix; we further ignore $\boldsymbol{H}$.

  It is not obvious which of the components $\boldsymbol{C}$ and $\boldsymbol{D}$ is responsible for the conversion from electromagnetic field to elecrical voltage. Even though one may argue that the conversion is part of the intended feed behaviour, the choice is actually quite arbitrary in that it has no consequences.
- The complex receiver gains, $\boldsymbol{G}$.

Where a permanent connection exists between the feed and the receiver system, the feed and gain matrices always appear in the combination $\boldsymbol{GD}$. We may therefore represent them together as a single component, the *receiver matrix*

$$\boldsymbol{R}_{\mathrm{A}} = \boldsymbol{G}_{\mathrm{A}}\boldsymbol{D}_{\mathrm{A}}$$

For describing the system, it is the effect of $\boldsymbol{R}$ that matters; the fact that it is the product of two other matrices is irrelevant. (A familiar analogue of this is to be found in the *phase* of an interferometer: One knows that it is the sum of many contributions, but for the interferometer as a whole only this sum matters.) In practical applications, for various reasons one is often interested in the separate contributions of $\boldsymbol{D}$ and $\boldsymbol{G}$ to system behaviour, so we will consider the two matrices separately where appropriate.

The equation describing the signal path in antenna A is then

$$\boldsymbol{v}_{\mathrm{A}} = \boldsymbol{J}_{\mathrm{A}}\boldsymbol{e}_{\mathrm{A}}\ ;\quad \boldsymbol{J}_{\mathrm{A}} = \boldsymbol{R}_{\mathrm{A}}\boldsymbol{C}_{\mathrm{A}}\boldsymbol{P}_{\mathrm{A}}\boldsymbol{F}_{\mathrm{A}} \tag{11}$$

From the identity established in Eqs. (5) and (6), it follows that for the interferometer we have a corresponding equation in the coherency domain; we add the Stokes transformation to get:

$$\boldsymbol{v} = \boldsymbol{JS}\ \boldsymbol{e}^{\mathrm{S}} = \boldsymbol{K}\ \boldsymbol{e}^{\mathrm{S}}\ ;\quad \boldsymbol{J} = \boldsymbol{J}_{\mathrm{A}} \otimes \boldsymbol{J}^{*}_{\mathrm{B}} \tag{12}$$

$\boldsymbol{K}$ gives the overall system response as observed with a radio interferometer; we call it the *system matrix*. $\boldsymbol{J}$ represents the part of $\boldsymbol{K}$ that can be factored; this factoring is the basis of the self-calibration methods to be discussed in Paper II.

$\boldsymbol{J}$ may be calculated as the outer product of two cumulative matrix products or vice versa. The former is simpler to evaluate because the matrices are only 2×2. The matrix $\boldsymbol{S}$ cannot be represented as an outer product, so the final multiplication is of two 4×4 matrices in the coherency domain. A technique to simplify this multiplication is shown in Appendix B.

Figure 3 is a schematic showing the paths of the signal and coherency vectors and the various coordinate systems in which they are described.

As an aside we note that, because of the non-decomposability of $\boldsymbol{S}$ and its inverse, it is impossible to design a combination of feeds that measures the Stokes visibilities directly.

4.2. *The individual Jones matrices*

We now consider the form of each of the matrices in Eq. (11).

Both the Faraday and the parallactic rotations $\boldsymbol{F}_{\mathrm{A}}$ and $\boldsymbol{P}_{\mathrm{A}}$ are simple rotations around the line of sight represented by rotation matrices

$$\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{13}$$

ϕ is positive for a Faraday rotation as indicated in Fig. 2; it is negative for a feed rotation in the same sense.

The configuration matrix $\boldsymbol{C}$ is defined by the *nominal* properties of the feed, i.e. the *intended* transformation from input e.m. field to output voltage. It includes

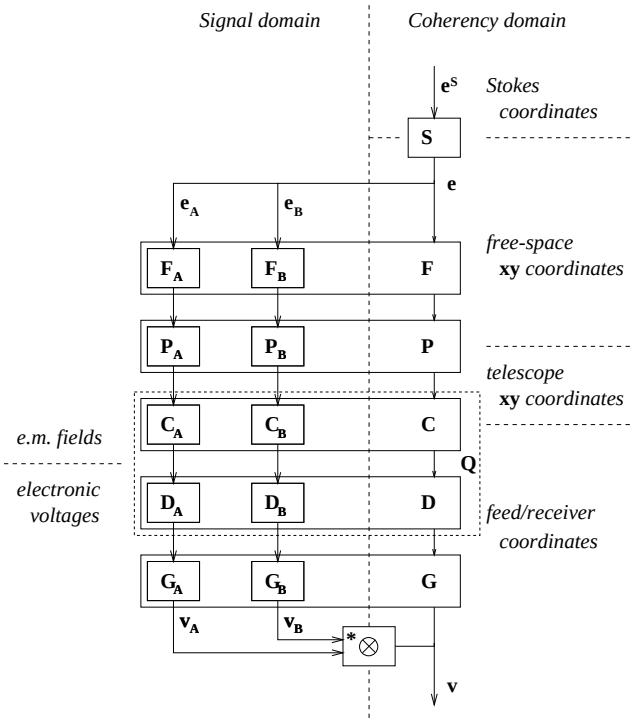

Fig. 3. Schematic showing the path of the signal and coherency vectors in their respective domains. To the right the coordinate systems are shown in which they are represented

both the effects of feed design (e.g. linearly or circularly polarized) and of any deliberate rotation.

The feed-error matrix $\boldsymbol{D}$ ('D' for 'dipole') represents the deviations of the *true* feed from its design. The general form we shall use is

$$\boldsymbol{D}_{\mathrm{A}} = \begin{pmatrix} 1 & d_{\mathrm{Ap}} \\ -d_{\mathrm{Aq}} & 1 \end{pmatrix} \tag{14}$$

where d_{Ap} represents the spurious sensitivity of the p receptor to the q polarization and vice versa. The justification of Eq. (14) is given in Appendix D.

The electronic gain is represented by a simple diagonal matrix

$$\boldsymbol{G}_{\mathrm{A}} = \begin{pmatrix} g_{\mathrm{Ap}} & 0 \\ 0 & g_{\mathrm{Aq}} \end{pmatrix} \; ; \; \boldsymbol{G} = \begin{pmatrix} g_{\mathrm{pp}} & 0 & 0 & 0 \\ 0 & g_{\mathrm{pq}} & 0 & 0 \\ 0 & 0 & g_{\mathrm{qp}} & 0 \\ 0 & 0 & 0 & g_{\mathrm{qq}} \end{pmatrix}$$

4.3. *Stokes parameters in an interferometer*

Figure 2 shows the coherency vector in our interferometer system represented in xy coordinates. One may equally well represent it in the Stokes frame and consider the propagation of the Stokes vector visibility through the system; in this representation, the various 4×4 matrices assume the form of the so-called Mueller matrices of optics (Azzam & Bashara 1987; Simmons & Guttman 1970). However, in generalizing from traditional optical applications to the case of an interferometer, one must completely abandon the customary interpretation of the Stokes parameters and the intuition based on it.

To begin with, Stokes visibilities may assume complex values. Moreover, the traditional notion that a signal cannot be more that 100% polarized is no longer valid. This may be demonstrated by a simple example. Consider an interferometer with Faraday rotations of α and β, respectively, in its two arms. Then in Stokes coordinates

$$\boldsymbol{F}^{\mathrm{S}} = \boldsymbol{S}^{-1}(\boldsymbol{F}_{\mathrm{A}} \otimes \boldsymbol{F}_{\mathrm{B}}^{*})\boldsymbol{S} =$$

$$\begin{pmatrix} \cos(\alpha-\beta) & 0 & 0 & i\sin(\alpha-\beta) \\ 0 & \cos(\alpha+\beta) & -\sin(\alpha+\beta) & 0 \\ 0 & \sin(\alpha+\beta) & \cos(\alpha+\beta) & 0 \\ i\sin(\alpha-\beta) & 0 & 0 & \cos(\alpha-\beta) \end{pmatrix}$$

If $\alpha = \pi/2$, $\beta = 0$, then an unpolarized input signal, $\boldsymbol{e}^{\mathrm{S}} = (1,0,0,0)$ will emerge as $(0,0,0,i)$.

Most if not all existing software packages for the processing of radio-polarimetric observations make the assumption of *weak polarisation*, i.e. $Q,\ U,\ V \ll I$. This assumption, along with that of small receptor errors, allows one to suppress higher-order terms in the interferometer equation and simplify the non-linear problem to a linear one, cf. Sect. 5 below. It is, however, fundamentally incorrect: Indeed, it is quite possible for a source that is weakly polarized in its brightness to appear strongly polarized in its visibilities on particular baselines or with particular feed configurations. Where this happens, the linearized interferometer equations will produce incorrect results. The impact of such errors on the images of the sky that are eventually produced needs to be investigated.

A spectacular example where the observed visibilities are much more strongly polarised than the source brightness is to be found in the recent observations of galactic-foreground Faraday rotation by Wieringa et al. (1993). In these observations, brightness polarizations well above 100% are seen; this may happen because the distribution of Stokes I is so smooth that only a small fraction of it is picked up by the interferometers.

5. Practical applications

The main reasons why the black-box formula of Morris et al. (1964) is unsatisfactory in practice are *a)* it does not clarify the contributions of individual components of the interferometer system to its overall behaviour, and *b)* it does not include the effect of instrumental errors (apart from the orientations and ellipticities of the receptors). In Eqs. (11) and (12) each system component retains its identity and can be modified to include errors. We discuss here a few examples of the use of our formalism.

5.1. *The number of free parameters*

The first point to note is that Eq. (12) contains the two matrices $\boldsymbol{J}_{\rm A}$ and $\boldsymbol{J}_{\rm B}$ that contain unknown errors. These two matrices between them contain 8 complex elements, so the number of complex parameters to be determined by calibration is at most 7; indeed the number is one less than 8 because we may arbitrarily define one element and scale the remaining ones accordingly. This result is analogous to a theorem on Mueller matrices in optics (Azzam & Bashara 1987; O'Neill 1963).

In practice, the number of relevant complex parameters may become less when certain approximations are introduced. It is customary to assume that the feed errors are small, i.e. the off-diagonal elements of $\boldsymbol{D}_{\rm A}$ and $\boldsymbol{D}_{\rm B}$ are $\ll 1$. In the limited context of the examples to be discussed below, we also assume that the Stokes visibilities are weakly polarized, ignoring for a moment the objections of Sect. 4.3.

5.2. *The output of an interferometer with parallel linear feeds*

Several radio telescopes consist of interferometers with parallel linear feeds. With the above approximations it is a simple exercise to work out the matrix product $\boldsymbol{K}^{++} = \boldsymbol{G}(\boldsymbol{D}_{\rm A} \otimes \boldsymbol{D}_{\rm B}^*)\boldsymbol{S}$, assuming no rotation effects. Dropping products of small terms, one obtains

$$\boldsymbol{K}^{++} \approx \frac{1}{2}\begin{pmatrix} g_{\rm pp} & g_{\rm pp} & 0 & 0 \\ g_{\rm pq}\Delta_{\rm pq} & 0 & g_{\rm pq} & ig_{\rm pq} \\ -g_{\rm qp}\Delta_{\rm qp} & 0 & g_{\rm qp} & -ig_{\rm qp} \\ g_{\rm qq} & -g_{\rm qq} & 0 & 0 \end{pmatrix} \tag{15}$$

where

$$\Delta_{\rm pq} = (d_{\rm Ap} - d_{\rm Bq}^*)\ ;\ \ \Delta_{\rm qp} = (d_{\rm Aq} - d_{\rm Bp}^*)$$

This result is equivalent to that given for the Westerbork telescope by Weiler & Raimond (1976). Note that the number of relevant complex parameters to be determined has been reduced from 7 to 6.

5.3. *The output of an interferometer with 'crossed' linear feeds*

The Westerbork Telescope has also been operated in a so-called 'crossed-dipole' mode (Weiler 1973). The salient feature of this mode is that each interferometer combines linear feeds differing by $\pi/4$ in position angle (Weiler 1973; Thompson et al. 1986). It is not difficult to work out the complete system matrix for this case, but we may take a shortcut: We first derive $\boldsymbol{K}$ for the error-free case; we then introduce the errors as perturbations and show that we can absorb them in the gain factors G.

We assume the feed of antenna B to be rotated:

$$\boldsymbol{C}_{\rm B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and find

$$\boldsymbol{K}^{+\times} \approx \frac{1}{2\sqrt{2}}\begin{pmatrix} g_{\rm pp} & g_{\rm pp} & -g_{\rm pp} & ig_{\rm pp} \\ g_{\rm pq} & g_{\rm pq} & g_{\rm pq} & -ig_{\rm pq} \\ -g_{\rm qp} & g_{\rm qp} & g_{\rm qp} & ig_{\rm qp} \\ g_{\rm qq} & -g_{\rm qq} & g_{\rm qq} & ig_{\rm qq} \end{pmatrix} \tag{16}$$

The errors enter as small increments to the elements of $\boldsymbol{D}$; it is readily verified that consequently each element k_{ij} is multiplied by some factor $(1+\epsilon_{ij})$:

$$k_{11} = g_{\rm pp}(1+\epsilon_{11})\ ,\ \ \text{etc.}$$

where the ϵ_{ij}s are combinations of off-diagonal elements from $\boldsymbol{D}_{\rm A}$ and $\boldsymbol{D}_{\rm B}$. Following Weiler (1973), we now introduce the approximation

$$\epsilon_{ij} = \epsilon_{1j}\ ,i,j = 1,2,3,4$$

The first-order small errors that we so introduce are reduced to a negligible level in the multiplication with Q, U, and V which we assume to be small as before. We may then absorb the factors ϵ_{1j} in the gain factors, $g_{\rm pp}$ etc., which brings us back to Eq. (16). This equation, however, now represents the system matrix *including* feed errors to first order. In this approximation, the number of relevant complex parameters for this configuration is only 4.

Weiler's result differs from ours in the signs of most of the matrix elements. This is because he assumes feed A to be in position angle $\pi/2$ rather than 0.

5.4. *The similarity between linear- and circular-feed interferometers*

There is a well-known similarity between the parallel-linear interferometer of Sect. 5.2 and one having two circularly polarized feeds: Their behaviour with respect to the Stokes visibilities is identical except for a permutation of Q, U and V.

Our formalism can be used to prove this in a simple and elegant way. We begin by noting that the linear and circular feeds differ by the presence of the linear-to-circular coordinate transformation

$$\boldsymbol{C}_{\rm A} = \boldsymbol{C}_{\rm B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \tag{17}$$

Again assuming no rotation effects, we transform $\boldsymbol{C}$ to the Stokes coordinate frame:

$$\boldsymbol{RCS} = \boldsymbol{RSC}^{\rm S}\ ;\ \ \boldsymbol{C}^{\rm S} = \boldsymbol{S}^{-1}\boldsymbol{CS}$$

(We have discussed the transformation of an *operator* in Sect. 3.3. Transforming a *coordinate transformation* has no obvious physical interpretation, but there is nothing mathematically to prevent us from making such a transformation when it suits us.) The result is

$$\boldsymbol{C}^{\rm S} = \boldsymbol{S}^{-1}(\boldsymbol{C}_{\rm A} \otimes \boldsymbol{C}_{\rm B}^*)\boldsymbol{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{18}$$

Post-multiplication with this matrix amounts to the permutation of columns already mentioned.

6. Limitations of the Jones formalism: Decorrelation and depolarization

In Sect. 3.1 we postulated our signals to be quasi-monochromatic without stating what this means. We now want to examine the foundations of the Jones formalism more precisely.

An electromagnetic signal propagating in space may be represented at some fixed point as an integral of monochromatic components over its bandwidth $\Delta\omega$:

$$\boldsymbol{\epsilon}(t) = \int_{\Delta\omega} \mathrm{d}\omega\ \boldsymbol{a}(\omega)\ \mathrm{e}^{i\omega t} \tag{19}$$

where $\boldsymbol{a}(\omega)$ is a vector with complex components. We may rewrite this equation in the form

$$\boldsymbol{\epsilon}(t) = \mathrm{e}^{i\omega_0 t} \int_{\Delta\omega} \mathrm{d}\omega\ \boldsymbol{a}(\omega)\ \mathrm{e}^{i(\omega-\omega_0)t} = \boldsymbol{e}(t)\ \mathrm{e}^{i\omega_0 t}\ ,$$

i.e. we represent it as a monochromatic signal with a varying amplitude. In principle, any signal can be represented in this way. According to the Nyquist sampling theorem, $\boldsymbol{e}$ is essentially constant over time intervals $\ll \pi/\Delta\omega$. Thus, $\boldsymbol{e}(t)$ will vary slowly with respect to the oscillation period if the relative bandwidth $2\Delta\omega/\omega \ll 1$; this is the common definition of a quasi-monochromatic signal (see e.g. Born & Wolf 1964).

Consider a linear optical element in the signal path represented by an operator $\boldsymbol{J}(\omega)$, converting the signal $\boldsymbol{\epsilon}(t)$ of Eq. (19) into

$$\boldsymbol{\epsilon}'(t) = \int_{\Delta\omega} \mathrm{d}\omega\ \boldsymbol{J}(\omega)\boldsymbol{a}(\omega)\ \mathrm{e}^{i\omega t} \tag{20}$$

If $\boldsymbol{J}$ is independent of ω (except for delays which we ignore), then

$$\boldsymbol{\epsilon}'(t) = \boldsymbol{J}\boldsymbol{\epsilon}(t)$$

and consequently

$$\boldsymbol{e}'(t) = \boldsymbol{J}\boldsymbol{e}(t)$$

We conclude that the Jones formalism is valid for arbitrarily large relative bandwidths, as long as the operators $\boldsymbol{J}$ are frequency-independent.

Many real systems do not conform to this restriction. It is, however, possible, to conceptually subdivide the frequency band of interest into a number of subbands that each individually satisfy the requirement; this is in fact what Eq. (20) does. To obtain the correlator output, we must then sum over these subbands; substituting Eqs. (6) from Sect. 3 and (20) above in Eq. (10), we get

$$\boldsymbol{v} = \\ < \textstyle\sum_{i,j}\ \boldsymbol{J}_{\mathrm{A}}(\omega_i) \otimes \boldsymbol{J}^*_{\mathrm{B}}(\omega_j)\ \boldsymbol{a}_{\mathrm{A}}(\omega_i) \otimes \boldsymbol{a}^*_{\mathrm{B}}(\omega_j)\mathrm{e}^{i(\omega_i-\omega_j)} > = \\ < \textstyle\sum_{i}\ \boldsymbol{J}_{\mathrm{A}}(\omega_i) \otimes \boldsymbol{J}^*_{\mathrm{B}}(\omega_i)\ \boldsymbol{a}_{\mathrm{A}}(\omega_i) \otimes \boldsymbol{a}^*_{\mathrm{B}}(\omega_i) >$$

Without investigating this sum in detail, we may qualitatively understand an essential property by pretending for a moment that the vector variables in it are scalars. It is then clear that $\boldsymbol{v}$ will be maximal if all subband products $J_{\mathrm{A}}J^*_{\mathrm{B}}\ a_{\mathrm{A}}a^*_{\mathrm{B}}$ add in phase. Failure to meet this results in a loss of signal which in scalar interferometer theory is known as *bandwidth decorrelation*.

In radio interferometers this effect is preempted by, firstly, eliminating differential delays between the signals e_{A} and e^*_{B} and, secondly, matching the frequency dependencies of J_{A} and J_{B} as closely as possible. Where this is not possible (for fundamental reasons beyond the scope of this paper), the division into subbands is performed litterally: The wide-band input signal is distributed over a number of parallel narrower-band channels, each with its own correlator.

Depolarization is the more general appearance that bandwidth decorrelation assumes in a vector theory of interferometry and image formation. It may result not only from the integration over frequency that we just discussed, but also from other integrations, e.g. over the cross-section of a beam of radiation or over a source of finite extent. Faraday depolarization due to finite spatial resolution is a familiar effect in radio astronomy, and in optics depolarizing optical elements are well known. Such elements cannot be described by a Jones matrix, but do have a 4×4 *Mueller matrix* which describes the transmission of the Stokes vector.

Qualitatively, the concept of depolarization is easy to understand. From the preceding discussion it is clear that components representable by Jones matrices will not suffer from it. For this reason, such components are also referred to as *non-depolarizing* ones. This term is, however, a dangerous one because it too strongly suggests the naive interpretation that in transmission the fractional polarization

$$\sqrt{(Q^2+U^2+V^2)/I^2}$$

does not decrease. That this interpretation is incorrect, can easily be demonstrated by a counter-example:

Consider the Jones matrix

$$\boldsymbol{J}_{\mathrm{A}} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

The corresponding Mueller matrix is

$$\boldsymbol{S}^{-1}\ (\boldsymbol{J}_{\mathrm{A}} \otimes \boldsymbol{J}^*_{\mathrm{A}})\ \boldsymbol{S} = \begin{pmatrix} 5 & -4 & 0 & 0 \\ -4 & 5 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

This element converts a partially polarized signal, $\boldsymbol{e}^{\mathrm{S}} = (5, 4, 0, 0)$, to an unpolarized one, $\boldsymbol{e}^{\mathrm{S}\prime} = (9, 0, 0, 0)$.

The correct definition of the term 'non-depolarizing' is given by Ditchburn (1976), who states that a Jones operator converts a fully polarized signal into another one; i.e. if for the input signal

$$I^2 - Q^2 - U^2 - V^2 = 0 \tag{21}$$

then the same is true for the output. A proof of this property is shown in Appendix E. In Paper II, we will follow Hovenier (1994) in using the term *pure* for the Mueller matrix of an optical element representable by a Jones matrix.

7. Conclusions

The formalism developed in this paper puts radio-interferometric polarimetry on a solid theoretical footing. It provides a coherent picture of what happens polarization-wise in a radio interferometer and allows inclusion of the various instrumental effects in a straightforward way. It bridges an existing gap between the theoretical methods of optical and radio polarimetry in a very satisfactory way.

Crossing this bridge, one may borrow well-established results from the optical theories and bring then to bear on radio interferometry. Rather than looking at the mechanisms inside an interferometer as we have done here, one may consider an aperture-synthesis instrument as a whole and consider its properties as an imaging instrument. The insights that this viewpoint provides into the fundamental limitations on calibration schemes for synthesis arrays will be the subject of our Paper II (Sault et al. 1996).

In this paper we have assumed definitions for several relevant entities in the theory. Selecting the correct signs in these definitions appears to be a problem that has not been addressed in a systematic and complete way so far. As a complement to the present work, we have studied the problem in depth; the results and a critical review of earlier writings on the problem will be presented in Paper III (Hamaker & Bregman 1996).

An important point noted in this paper is that weak polarization in the brightness of a source does not necessarily imply the same for its observed visibilities. This point seems to have been overlooked heretofore, and its impact on the data reduction procedures that are routinely used in aperture synthesis should be investigated.

Acknowledgements. We acknowledge earlier work by J.E. Noordam and W.N. Brouw on which the present work is based and the stimulating interest of J. Tinbergen in its further development, as well as constructive criticism of the manuscript by A. van Ardenne, W.N. Brouw and J.W. Hovenier. The Netherlands Foundation for Research in Astronomy (NFRA) is operated with financial support from the Netherlands Organisation for Scientific Research (NWO). RJS acknowledges support from this organisation under their visiting scientist scheme.

A. The outer matrix product

A discussion of outer (also known as direct, tensor or Kronecker) products can be found in medium- and advanced-level matrix or linear algebra texts. This product, $\boldsymbol{A} \otimes \boldsymbol{B}$, is defined as a new matrix in which each element a_{ij} of $\boldsymbol{A}$ is replaced by $a_{ij}\boldsymbol{B}$. Thus the outer product of 2-element column vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is

$$\boldsymbol{a} \otimes \boldsymbol{b} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \\ a_2 b_1 \\ a_2 b_2 \end{pmatrix}$$

and the outer product of two 2×2 matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}$$

To us the most important property of the outer product is the redistribution relation Eq. (5). From it, a number of other properties can be proven by elementary manipulations:

- The transpose (inverse) of an outer product is the outer product of the transposes (inverses) of the factors.
- If the factors of an outer product are symmetric (hermitian, unitary), so is their outer product.

Note that our definition is based on a particular representation of the product's factors. It is not obvious that the product is invariant under coordinate transformations. It is easy to show, though, that if we transform the factors by transformations $\boldsymbol{T}_{\mathrm{A}}$ and $\boldsymbol{T}_{\mathrm{B}}$, respectively, the product is transformed by $\boldsymbol{T} = \boldsymbol{T}_{\mathrm{A}} \otimes \boldsymbol{T}_{\mathrm{B}}$. We satisfy ourselves with this simple observation, knowing that a proper axiomatic definition exists in the theory of tensor algebra. See e.g. Korn & Korn (1961) or Pipes & Harvill (1958) for a very formal introduction; for a more tutorial treatment one must refer to an appropriate textbook, e.g. Myškis (1975).

B. A decomposition of matrix $\boldsymbol{S}$

In computing the system matrix, it is advantageous to work in the signal domain as much as possible, because the matrices to be multiplied are only 2×2 there. Even the multiplication by $\boldsymbol{S}$ can be circumvented by writing $\boldsymbol{S}$ as the sum of two outer products, multiplied with a trivial 4×4 matrix

$$\boldsymbol{S} = \boldsymbol{S}' \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -i \end{pmatrix}$$

$$\begin{aligned}\boldsymbol{S}' &= \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \otimes \begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \otimes \begin{pmatrix}0 & 0\\ 1 & -1\end{pmatrix}\\ &= \boldsymbol{I} \otimes \boldsymbol{X} + \boldsymbol{J} \otimes \boldsymbol{Y}\end{aligned}$$

This allows us to cast the multiplication by $\boldsymbol{S}'$ of an outer product $\boldsymbol{W}_{\mathrm{A}} \otimes \boldsymbol{W}_{\mathrm{B}}^*$ into a form that is both easier for hand calculations and more economic in computer evaluation:

$$\begin{aligned}(\boldsymbol{W}_{\mathrm{A}} \otimes \boldsymbol{W}_{\mathrm{B}}^*)\, \boldsymbol{S}' &= (\boldsymbol{W}_{\mathrm{A}}\boldsymbol{I}) \otimes (\boldsymbol{W}_{\mathrm{B}}^*\boldsymbol{X})\\ &+ (\boldsymbol{W}_{\mathrm{A}}\boldsymbol{J}) \otimes (\boldsymbol{W}_{\mathrm{B}}^*\boldsymbol{Y})\end{aligned}$$

Another merit of this expression is that it puts into evidence certain symmetries to be expected in a system matrix.

C. Stokes parameters and Stokes visibilities

We briefly outline the concept of Stokes visibilities here; more thorough treatments exist in the literature (Conway & Kronberg 1969; Thompson et al. 1986).

Classical interferometry considering brightness alone is based on the van Cittert-Zernike theorem, which states that the spatial autocorrelation function of the electromagnetic field is the Fourier transform of the brightness distribution. This function is known as the *visibility* $V(\boldsymbol{r})$; its argument is the vector separation between the two sampling points.

Extending this concept to include polarization, one recognizes that the visibility function is now a vector entity whose four components coincide with those of the coherency vector in our paper. In other words, we identify the coherency vector $\boldsymbol{e}^+$ measured in geometric xy coordinates with the value of the vector visibility function when we take for its argument the vector separation between antennas A and B.

The classical definition of the Stokes parameters refers to the electric field at a single point in the radiation field, i.e. $\boldsymbol{r} = 0$. The generalization to include the dependence on $\boldsymbol{r}$ is obvious and leads to the concept of the *Stokes visibility functions*:

$$\boldsymbol{e}^{\mathrm{S}}(\boldsymbol{r}) = \boldsymbol{S}\boldsymbol{e}^+(\boldsymbol{r})$$

Unlike the classical parameters which are real, the Stokes visibilities are *complex* functions. It is readily apparent that they are Hermitian, i.e.

$$\boldsymbol{e}(\boldsymbol{r}) = \boldsymbol{e}^*(-\boldsymbol{r})$$

Fourier-transforming this function back to sky brightnesses, we get the *real* brightness distributions in each of the four Stokes parameters. Since both the Stokes and Fourier transformations are linear, one could also first Fourier transform the vector visibility function and apply the Stokes transformation afterwards for each point on the sky.

D. The feed model

The feed consists of two receptors. Each is designed to be sensitive to only one of two mutually *opposite* polarizations. Defining the coordinate axes in which we represent the incident radiation to coincide with these two polarizations, the feed would ideally be represented by a unit matrix. In practice, there is some *leakage* of the opposite polarization into either receptor, so the actual feed matrix for antenna A is

$$\begin{pmatrix} D'_{\mathrm{Ap}} & d'_{\mathrm{Ap}}\\ -d'_{\mathrm{Aq}} & D'_{\mathrm{Aq}}\end{pmatrix}$$

The two rows of this matrix represent the p and q receptors, respectively. For a well-designed feed, the off-diagonal *leakage* terms are small and the diagonal terms are close to unity.

We decompose this matrix, writing d_{Ap} for $d'_{\mathrm{Ap}}/D'_{\mathrm{Ap}}$, to get

$$\begin{pmatrix} D'_{\mathrm{Ap}} & 0\\ 0 & D'_{\mathrm{Aq}}\end{pmatrix}\begin{pmatrix} 1 & d_{\mathrm{Ap}}\\ -d_{\mathrm{Aq}} & 1\end{pmatrix} = \boldsymbol{G}'\boldsymbol{D}$$

Of these two matrices, $\boldsymbol{G}'$ represents a gain factor that we may absorb in the gain matrix $\boldsymbol{G}$. The feed characteristics are then adequately represented by $\boldsymbol{D}$, cf. Eq. (14).

It is important to realize that this model depends on first principles only and does not involve any approximations. Some references start from the specific model that describes a receptor in terms of its *orientation* and *ellipticity* (Born & Wolf 1964; Thompson et al. 1986), suggesting that Eq. (14) is valid only in first order if the errors in these quantities are small. This assertion results from failing to recognize the possibility of factoring off the gain term, $\boldsymbol{G}'$. See Sault et al. (1991) for a correct treatment of the orientation/ellipticity model in accordance with our description.

The first-order approximation is often helpful, however, by associating the real and imaginary parts of a d term with orientation and ellipticity errors, respectively. However, effects other than just orientation and ellipticity errors may also produce leakage.

D.1. *Linear feeds with a linear-to-circular converter*

In Sect. 4.1 we state that the combination of a linearly polarized feed and linear-to-circular hybrid converter can be represented in the more desirable form of a configuration and a feed matrix. This is important because conceptually the converter is part of the configuration and we do not want two configuration matrices.

Let the ideal hybrid be represented by $\boldsymbol{H}$ and the real one by $\boldsymbol{H}'$. Then

$$\boldsymbol{H}'\boldsymbol{D}\boldsymbol{C} = (\boldsymbol{H}'\boldsymbol{D}\boldsymbol{H}^{-1})\,(\boldsymbol{H}\boldsymbol{C}) = \boldsymbol{D}'\boldsymbol{C}'$$

The new configuration matrix $\boldsymbol{C}'$ now includes the transformation to circular-rl coordinates and thus describes a

fiducial, nominally circularly polarized feed; $\boldsymbol{D}'$ describes the errors of this fiducial feed.

E. Pure Mueller matrices

That a Jones matrix is non-depolarizing in the sense defined by Ditchburn (1976), cf. Sect. 6), is easily shown. We start by noting that, for a fully polarized signal (Azzam & Bashara 1987)

$$e_x/e_y = c \ , \quad \text{a constant} \tag{E1}$$

The value of c determines the type of polarization, e.g. for $c = 0$ we have a signal linearly polarized in the y direction, for $c = \pm i$ we have a left- or right-circularly polarized signal, etc. The Jones matrix converts this constant ratio into another ratio which is again a constant, i.e. the output signal is again fully, albeit in general differently, polarized.

To see the relation to Eq. (21), we transform that equation to the geometric xy coordinate frame by substituting Eq. (8), which yields a corresponding relation between the xy components of the coherency vector. We express this relation in terms of signal-vector components in order to explicitly show the averaging involved:

$$< e_{\mathrm{A}x} e^*_{\mathrm{A}x} >< e_{\mathrm{A}y} e^*_{\mathrm{A}y} > - < e_{\mathrm{A}x} e^*_{\mathrm{A}y} >< e_{\mathrm{A}y} e^*_{\mathrm{A}x} >= 0 \tag{E2}$$

It is clear that condition Eq. (E1) is sufficient for this condition to hold. It is also necessary: Indeed, Eq. (E2) represents the boundary case of a Cauchy-Schwarz inequality; the required equality exists only if Eq. (E1) holds (Korn & Korn 1961).

References

Azzam R.M.A., Bashara N.M., 1987, Ellipsometry and polarized light. North Holland Physics Publishing, Amsterdam

Born M., Wolf E., 1964, Principles of Optics. Pergamon Press

Conway R.G., Kronberg P.P., 1969, MNRAS 142, 11-32

Ditchburn R.W., 1976, Light, Vol. I. Academic Press

Hamaker J.P., Bregman J.D., 1996, A&AS 117, 161

Hecht E., Zajac A., 1982, Optics Addison-Wesley

Hovenier J, 1994, Appl. Opt. 33, 8318-8324

IAU, 1973, Trans. IAU 15b, 166

Jones R.C., 1941, J. Opt. Soc. America 31, 488-493

Korn G.A., Korn T.M., 1961, Mathematical Handbook for Scientists and Engineers. McGraw Hill

Morris D., Radhakrishnan V., Seielstad G.A., 1964, ApJ 139, 551-559, The derivation of the black-box formula is reproduced in Weiler (1973) and Thompson et al. (1986)

Mueller H., 1948, J. Opt. Soc. America 38, 661, see also many textbooks on Optics, e.g. Azzam & Bashara (1987), Hecht & Zajac (1982)

Myškis A.D., 1975, Advanced Mathematics for Engineers (translated from Russian). Mir Publishers, Moscow

O'Neill E.L., 1963, Introduction to Statistical Optics. Addison-Wesley

Pipes L.A., Harvill L.R., 1958, Applied Mathematics for Engineers and Physicists. McGraw Hill

Sault R.J., Killeen N.E.B., Kesteven M.J., 1991, AT Technical Document Series 39.3/015

Sault R.J., Hamaker J.P., Bregman J.D., 1996, A&AS 117, 149

Simmons J.W., Guttman M.J., 1970, States, waves and photons: A modern introduction to light. Addison-Wesley

Stokes G.G., 1852, Trans. Cambridge Phil. Soc. 9, part III, 399-416. Reprinted in: 1901 Mathematical and Physical Papers 3, 233-258. Cambridge Univ. Press

Thompson A.R., Moran J.M., Swenson G.W. Jr, 1986, Interferometry and Synthesis in Radio Astronomy. John Wiley & Sons, New York

van de Hulst H.C., 1957, Light scattering by small particles. Wiley, New York, 1957. Reprinted 1981, Dover, New York

Weiler K.W., 1973, A&A 26, 404-407

Weiler K.W., Raimond E., 1976, A&A 52, 397-402

Wieringa M.H., de Bruyn A.G., Jansen D., Brouw W.N., Katgert P., 1993, A&A 268, 215-229