Handling Polarized Light

Mark Shtaif

School of Electrical Engineering, Dept. of Physical Electron. Tel-Aviv University, Tel-Aviv, Israel shtaif@eng.tau.ac.il

This is supplementary material to three lectures that I am giving about polarization in my graduate course in fall 2004/5. The description and notations used are mostly consistent with those of Gordon and Kogelnik[1].

1 The Jones calculus for handling polarized light

In principle, the polarization of light is the direction of the electric fields vector of the light signal at any position and time. In cases of relevance it assumes a simpler and more useful interpretation. Recall that in single mode fibers the electric field vector is given by

$$\vec{\mathcal{E}}(\vec{r},t) = F(x,y)\vec{E}(t,z)e^{i\beta_0 z - i\omega_0 t}. \tag{1}$$

As we discussed earlier $F(x,y)$ is nearly independent of the things that matter in propagation and we may ignore it in what follows. We also assume that the instantaneous polarization (direction) of $\vec{E}(t,z)$ is independent of the position inside the beam (namely, it is independent of the x,y coordinates) thereby implying that we are treating the propagating field as a plane wave. This is known to be an excellent approximation in single mode fibers.

Let us now arbitrarily define the two orthogonal directions $\hat{e}_x$ and $\hat{e}_y$. The instantaneous field can then be expressed as

[1] J.P. Gordon and H. Kogelnik, "PMD fundamentals" in Proceedings of National Academy of Science. USA **97**, 4541 (2000), available on the web at http://www.pnas.org/cgi/content/abstract/97/9/4541.

$$\vec{E}_{Inst}(t,z) = \text{Real}\left(\vec{E}(t,z)e^{i\beta_0 z - i\omega_0 t}\right)$$
$$= |E_x|\cos(\omega_0 t - \beta_0 z - \varphi_x)\hat{e}_x + |E_y|\cos(\omega_0 t - \beta_0 z - \varphi_y)\hat{e}_y \quad (2)$$

where $E_x = |E_x|\exp(\imath\varphi_x)$ and $E_y = |E_y|\exp(\imath\varphi_y)$ are the complex amplitudes of the x and y components of the field, respectively. If we fix the value of z to 0 (without loss of generality) and plot the $\vec{E}(t,0)$ as a function of t we obtain an ellipse in the x, y plane. Thus a general state of polarization of such a plane wave sort of field is elliptic. There are two special cases of interest. One is when the two components along $\hat{e}_x$ and $\hat{e}_y$ are in phase with each other, namely when $\varphi_x = \varphi_y$, or $\varphi_x = \pi + \varphi_y$. In this case the width of the ellipse reaches 0 and we have oscillations of the field along some line in the x, y plane. This is called a *linear* state of polarization. The above equation in this case assumes the form

$$\vec{E}_{Inst}(t) = (|E_x|\hat{e}_x \pm |E_y|\hat{e}_y)\cos(\omega_0 t - \varphi) \quad (3)$$

where we have omitted the dependence on z, for simplicity. The other spacial case is when $\varphi_x - \varphi_y = \pm\pi/2$ and $|E_x| = |E_y|$. In that case the above can be written as

$$\vec{E}_{Inst}(t) = |E_x|\cos(\omega_0 t - \varphi_x)\hat{e}_x \pm |E_y|\sin(\omega_0 t - \varphi_x)\hat{e}_y \quad (4)$$

If the sign before the sine function is minus then the vector rotates clockwise and we have a right circular polarization. Otherwise it rotates anti-clockwise and the polarization is left circular.

All polarizations that are neither circular, or linear are generically called elliptic, and each elliptic polarization can have a right or left handedness as we shall have a chance to see in what follows. In general, the term *state of polarization* (SOP) of an optical field is defined in terms of the ellipse that is drawn by the electric field vector in the x, y plane. It is important to understand that the SOP has no dependence at all on the phase of the optical field, or on its intensity. It does however strongly depend on the phase difference between the field components and on their amplitude ratio.

The Jones representation- A convenient way of representing polarization vectors is with a column two dimensional vector

$$\begin{pmatrix} E_x(t) \\ E_y(t) \end{pmatrix} \tag{5}$$

whose two components are the complex envelopes of the signal along the $\hat{e}_x$ and $\hat{e}_y$ polarizations. Notice however that there is nothing special about the x, y directions and we can express the Jones vector in any other basis of orthogonal directions $\hat{e}_1$ and $\hat{e}_2$

$$\begin{pmatrix} E_1(t) \\ E_2(t) \end{pmatrix}_{\hat{e}_{1,2}} \tag{6}$$

This representation is named after Jones and the vector expressed in this form is called the Jones vector. The index $\hat{e}_{1,2}$ is included for clarity to indicate the basis in which the Jones vector is written. Obviously the actual $\vec{E}(t, z)$ is given by

$$\vec{E}(t) = E_1(t)\hat{e}_1 + E_2(t)\hat{e}_2. \tag{7}$$

Dot-products and Orthogonality- One may define orthogonality of polarization states as the geometrical orthogonality of two vectors (like $\hat{e}_x$ and $\hat{e}_y$). This may appear natural, but this is really not a satisfactory definition. Let us consider the dot product between two vectors $\vec{U} = [U_1, U_2]^t$ and $\vec{V} = [V_1, V_2]^t$. For simplicity we will assume that the envelopes $V_{1,2}$ and $U_{1,2}$ are independent of t, i.e. the CW case. Writing these two quantities explicitly we have

$$\begin{aligned} & [|U_1| \cos(\omega_0 t - \varphi_{U,1})\hat{e}_1 + |U_2| \cos(\omega_0 t - \varphi_{U,2})\hat{e}_2] \\ \cdot \; & [|V_1| \cos(\omega_0 t - \varphi_{V,1})\hat{e}_1 + |V_2| \cos(\omega_0 t - \varphi_{V,2})\hat{e}_2] \\ = \; & |U_1 V_1| \cos(\varphi_{U,1} - \varphi_{V,1}) + |U_2 V_2| \cos(\varphi_{U,2} - \varphi_{V,2}) \end{aligned}$$

where I have used simple trigonometric relations and removed terms that oscillate at optical frequencies. It is rather obvious that the above dot product can be re-expressed as

$$\mathrm{Real}(\vec{U}^\dagger \vec{V}) = \mathrm{Real}\left[\begin{pmatrix} U_1^* \; U_2^* \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix}\right]. \tag{8}$$

If the dot product between two vectors is 0 then the vectors are orthogonal. Notice that there are two ways in which the vector fields end-up being orthogonal. One is if the product $\vec{U}^\dagger \vec{V} = 0$. This is what we define as *orthogonality of polarizations*. The other is when the quantity $\vec{U}^\dagger \vec{V}$ is nonzero, but purely imaginary such that its real part is 0. In the latter case the polarizations are NOT orthogonal (recall that our definition of SOP does not depend on the phase), but the fields are said to belong to orthogonal quadratures. For example, the fields $(1, i)^t$ and $(i, 1)^t$ are orthogonal in polarization, and the fields $(1, i)^t$ and $(0, 1)^t$ are not orthogonal in polarization, but they are orthogonal because they belong to different quadratures. In the context of this class (dealing with polarizations) we will only be concerned with orthogonality in the polarization sense. In general the quantity $\vec{U}^\dagger \vec{V}$ will be referred to as the dot product between two Jones vectors.

Lets consider now the representation of a number of polarizations. The linear polarization case can always be expressed in the form

$$U_{\mathrm{linear}} = \exp(i\varphi)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{9}$$

with $a_{1,2}$ being real. A right and left circular polarizations are represented by the Jones vectors

$$\hat{e}_L = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 1 \end{pmatrix}, \qquad \hat{e}_R = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{10}$$

Of course we may also multiply them by any phase factor without changing anything (e.g. the Jones vector $(1, i)^t/\sqrt{2}$ also represents L.C. polarization, and $(i, 1)^t/\sqrt{2}$ is again a R.C. polarized unit vector).

Representation with different bases - Notice now that when we choose a basis $\hat{e}_{1,2}$ in which we wish to represent our polarization vector we are free to choose generic polarization states such as, for example, the right and left circular polarizations, that are obviously orthogonal. Here is an exercise; Suppose that a vector $\vec{U}$ is represented as

$$\vec{U} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}_{\hat{e}_{x,y}} \tag{11}$$

and we wish to represent it in the basis of RC and LC polarizations. Note that we may write $\hat{e}_x = \frac{1}{\sqrt{2}}(\hat{e}_R + i\hat{e}_L)$ and $\hat{e}_y = \frac{1}{\sqrt{2}}(i\hat{e}_R + \hat{e}_L)$. Thus, we do the following

$$\begin{aligned} \vec{U} &= u_x\hat{e}_x + u_y\hat{e}_y \\ &= u_x(\hat{e}_R + i\hat{e}_L)/\sqrt{2} + u_y(i\hat{e}_R + \hat{e}_L)/\sqrt{2} \\ &= (u_x + iu_y)/\sqrt{2}\hat{e}_R + (iu_x + u_y)/\sqrt{2}\hat{e}_L \end{aligned}$$

Or we may also write

$$\vec{U} = \begin{pmatrix} u_R \\ u_L \end{pmatrix}_{\hat{e}_{R,L}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}_{\hat{e}_{x,y}} \tag{12}$$

Notice that the conversion matrix from the $\hat{e}_x, \hat{e}_y$ basis to the $\hat{e}_R, \hat{e}_L$ basis, is unitary and its columns are the vectors $\hat{e}_x, \hat{e}_y$ represented in the $\hat{e}_R, \hat{e}_L$ basis according to what I wrote above. The matrix that translates the representation in the opposite direction (from the $\hat{e}_R, \hat{e}_L$ basis to the $\hat{e}_x, \hat{e}_y$ basis) is the conjugate of the above matrix and its columns are the vectors $\hat{e}_R, \hat{e}_L$, represented in the $\hat{e}_x, \hat{e}_y$ basis according to (10). Since the definition of polarization is not unique (it can be multiplied by any phase), the transformation matrix is also not unique.

The definition of a polarized signal- Returning to the case where the complex envelope is time dependent, so that we have a vector $\vec{U}(t)$ describing our signal. We may say that our signal is *polarized* if $\vec{U}(t)$ can be expressed as $\vec{U}(t) = U(t)\hat{u}$ such that $U(t)$ is scalar and $\hat{u}$ is time independent. At any moment in time (for the complex envelope) or at any given frequency all fields are perfectly polarized.

Transmission of polarized light through linear optical elements- The generalization of the scalar complex transfer function relating complex envelopes at the input and output of a linear element is a 2 by 2 frequency dependent matrix called the Jones matrix. Thus we may write

$$\vec{U}_{out}(\omega) = \boldsymbol{T}(\omega)\vec{U}_{in}(\omega) \tag{13}$$

where $\boldsymbol{T}$ denotes the Jones matrix. The combined matrix of cascaded elements is equal to the ordered product of the individual matrices. When the optical element is lossless, the matrix needs to be unitary because

$$|U_{out}|^2 = \vec{U}_{in}^{\dagger}\boldsymbol{T}^{\dagger}\boldsymbol{T}\vec{U}_{in} = |U_{in}|^2 \tag{14}$$

So that $\boldsymbol{T}^{\dagger}\boldsymbol{T} = \mathbf{1}$. We may also write that the matrix $\boldsymbol{T}$ must have the form

$$\exp(i\theta)\begin{pmatrix} r & t^* \\ t & -r^* \end{pmatrix} \tag{15}$$

where $|r|^2 + |t|^2 = 1$. In these lectures we will focus only on unitary elements. In reality the unitarity of the problem does not necessitate the assumption that these elements are truly lossless. It is enough to assume that they possess no polarization dependent loss, because polarization independent loss does not affect the polarization and can be factored out easily without influencing our analysis. A large fraction of optical components (including fibers) is unitary to a very good extent. Unitary matrices have a number of relevant properties. Most importantly, their eigenvectors are orthogonal to each other and the absolute value of both their eigenvalues is 1. Consequently, the eigenvectors of a unitary matrix form a basis in the complex 2 dimensional Jones space and if the matrix is represented in the basis of its eigenvectors it assumes the form

$$\boldsymbol{T} = \exp(i\bar{\varphi})\begin{pmatrix} \exp(i\Delta\varphi/2) & 0 \\ 0 & \exp(-i\Delta\varphi/2) \end{pmatrix}. \tag{16}$$

In the context of this section this implies that every unitary medium has two orthogonal SOPs that are eigenvectors of the medium and therefore pass through it unchanged. All other SOPs can be considered as a superposition of their components along the eigenvectors of the medium. The components aquire a phase difference of $\Delta\varphi$ between them as they propagate, thereby changing the SOP. Finally, I should emphasize that the eigenvectors and eigenvalues may all depend on the optical frequency ω. This means that in the above representation the basis vectors themselves are frequency dependent so that the above way of expressing the matrix $\boldsymbol{T}$ in the general case is useful only if we are considering a single optical frequency.

A word on Notation-This is a good point to introduce a more efficient notation for Jones analysis. We will leave the notation of Jones matrices as is, namely we will denote them with boldface letters. For the vectors on the other hand we adopt the Bra Ket notation. Thus a Jones vector will be denoted with a Ket $\vec{U} \to |U\rangle$ and the hermitian conjugate vector is denoted with a Bra $\vec{U}^\dagger \to \langle U|$. The inner product between $\vec{U}$ and $\vec{V}$ is now represented by $\vec{U}^\dagger\vec{V} \to \langle U|V\rangle$ and so on.

2 Introducing the Stokes representation

I start with a short mathematical detour whose pupose will be appreciated later. First, let me introduce the Pauli matrices in the context of polarizations

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \boldsymbol{\sigma}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \boldsymbol{\sigma}_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{17}$$

These are the same Pauli matrices known from the quantum theory of two-level systems except that they are permuted in order to yield results that are consistent with what is customary in the polarization jargon. These matrices together with the 2 by 2 identity matrix can be viewed as a basis for the space of 2 by 2 Hermitian matrices. This means that any hermitian matrix $\boldsymbol{H}$ can be represented as

$$\boldsymbol{H} = h_0\mathbf{1} + \sum_{j=1}^{3} h_j \boldsymbol{\sigma}_j \tag{18}$$

where the coefficients are all real numbers and $\mathbf{1}$ is the 2 by 2 identity matrix. A more useful notation is

$$H = h_0\mathbf{1} + \vec{h} \cdot \vec{\boldsymbol{\sigma}} \tag{19}$$

where $\vec{\boldsymbol{\sigma}} = [\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3]$ is a vector of matrices. Notice that I am using boldface letters to indicate 2 by 2 matrices or operators in Jones space. The pauli matrices have a number of relevant properties. Here is the first

$$\boldsymbol{\sigma}_j^2 = I, \qquad \boldsymbol{\sigma}_j \boldsymbol{\sigma}_k = i\boldsymbol{\sigma}_l \tag{20}$$

with (j, k, l) being any cyclic permutation of $(1, 2, 3)$. Another relevant property, which results trivially from (20) is that the Pauli matrices are trace orthogonal. i.e.

$$\text{Trace}\left[\boldsymbol{\sigma}_j \boldsymbol{\sigma}_k\right] = 2\delta_{j,k} \tag{21}$$

Relation (20) also implies that any function of the pauli matrices that can be expressed in a power series can be reduced to a have a linear dependence on the pauli matrices and the identity matrix $\mathbf{1}$. A number of additional properties to be proved by you are

$$\vec{\boldsymbol{\sigma}}(\vec{a} \cdot \vec{\boldsymbol{\sigma}}) = \vec{a}\mathbf{1} + i\vec{a} \times \vec{\boldsymbol{\sigma}} \tag{22}$$

$$(\vec{a} \cdot \vec{\boldsymbol{\sigma}})\vec{\boldsymbol{\sigma}} = \vec{a}\mathbf{1} - i\vec{a} \times \vec{\boldsymbol{\sigma}} \tag{23}$$

$$(\vec{a} \cdot \vec{\boldsymbol{\sigma}})(\vec{b} \cdot \vec{\boldsymbol{\sigma}}) = (\vec{a} \cdot \vec{b})\mathbf{1} + i(\vec{a} \times \vec{b}) \cdot \vec{\boldsymbol{\sigma}} \tag{24}$$

$$(\vec{a} \cdot \vec{\boldsymbol{\sigma}})\vec{\boldsymbol{\sigma}}(\vec{b} \cdot \vec{\boldsymbol{\sigma}}) = 2\vec{a}(\vec{a} \cdot \vec{\sigma}) - a^2\vec{\sigma} \tag{25}$$

$$\exp(i\vec{a} \cdot \vec{\boldsymbol{\sigma}}) = \cos(a)\mathbf{1} + i\sin(a)\hat{a} \cdot \vec{\sigma} \tag{26}$$

with a being the modulus of $\vec{a}$.

Now lets look at polarizations again. As we discussed earlier, a Jones vector $|U\rangle$ is characterized by a given energy (or power) represented by its square modulus $|U|^2 = \langle U|U\rangle$, by a certain phase and by its state of polarization. We already explained that the state of polarization is independent of neither the power of the signal nor of its phase. Let us now introduce a new creature, which we will refer to as the coherency matrix of the vector $|U\rangle$ and is expressed as $|U\rangle\langle U|$. In a column vector representation by some basis this can be written as a 2 by 2 matrix

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \begin{pmatrix} u_1^* & u_2^* \end{pmatrix} = \begin{pmatrix} |u_1|^2 & u_1 u_2^* \\ u_1^* u_2 & |u_2|^2 \end{pmatrix} \tag{27}$$

Notice that $|U\rangle\langle U|$ contains all the information regarding the power of $|U\rangle$ and its state of polarization. Only the phase of $|U\rangle$ cannot be retrieved. Also notice that $|U\rangle\langle U|$ is by definition a Hermitian matrix and as such there must exist a real vector $\vec{S}$ such that

$$|U\rangle\langle U| = \frac{1}{2}S_0\mathbf{1} + \frac{1}{2}\vec{S}\cdot\vec{\boldsymbol{\sigma}}, \tag{28}$$

where the 1/2 factor was introduced for later convenience. Since the pauli matrices are traceless it is obvious that S_0 is equal to the trace of $|U\rangle\langle U|$, namely $S_0 = |U|^2$. Also, noting that $|U\rangle\langle U||U\rangle\langle U| = |U|^2|U\rangle\langle U|$ and making use of relation (24) it can easily be shown (show it) that $|\vec{S}| = S_0 = |U|^2$. Consequently the subscript "0" will be dropped from S_0 and S will be used to denote both S_0 and the modulus of $\vec{S}$ in what follows. An alternative way of expressing the above is the following

$$|U\rangle\langle U| = \frac{|U|^2}{2}\left(\mathbf{1} + \hat{s}\cdot\vec{\boldsymbol{\sigma}}\right) \tag{29}$$

where $\hat{s} = \vec{S}/S$ is a unit vector indicating the orientation of $\vec{S}$. In the jargon of polarizations $\vec{S}$ is called the Stokes vector after its inventor[2] and $\hat{s}$ is the

[2] G.G. Stokes introduced the components of the Stokes vector in Trans. Cambridge Philos. Soc. 9 (1852) p. 399. He described the vector in terms of its components expressed in a way that is similar to equations (31)-(33). The formalism using the Pauli matrices was introduced much later.

normalized Stokes vector. Equation (29) tells us a number of things about the vector $\hat{s}$. First, we see that $|U\rangle$ is an eigenvector of $\hat{s}\cdot\vec{\boldsymbol{\sigma}}$ with the eigenvalue 1. Secondly, since $|U\rangle\langle U|$ is a Hermitian operator it has two eigenvectors that are orthogonal to each other. This implies that the other eigenvector of $\hat{s}\cdot\vec{\boldsymbol{\sigma}}$ is the Jones vector $|U_\perp\rangle$ that is orthogonal to $|U\rangle$ (namely $\langle U|U_\perp\rangle = 0$). The other eigenvalue of $|U\rangle\langle U|$, the one that corresponds to $|U_\perp\rangle$ is obviously 0 implying that $|U_\perp\rangle$ is also an eigenvector of $\hat{s}\cdot\vec{\boldsymbol{\sigma}}$ with eigenvalue -1.

It is now absolutely clear that the normalized Stokes vector $\hat{s}$ characterizes the SOP of our vector $|U\rangle$. Jones vectors with different SOPs will necessarily have different Stokes vectors in 3 dimensional space that correspond to them and Jones vectors with orthogonal SOPs will have anti-parallel Stokes vectors corresponding to them. If we look at the normalized Stokes vectors corresponding to all possible polarizations, their tips lie on a unit sphere called the Poincare' sphere. Every point on the sphere represents a unique SOP. Antipodal points on the sphere represent orthogonal SOPs. Now suppose that we have a vector $|U\rangle$. How do we extract the components of the Stokes vector $\vec{S}$? The best procedure is easily found from relation (21). Multiplying both sides of (29) by $\boldsymbol{\sigma}_j$ and taking the trace we find that

$$\mathrm{Trace}\big(|U\rangle\langle U|\boldsymbol{\sigma}_j\big) = \langle U|\boldsymbol{\sigma}_j|U\rangle = |U|^2\hat{s}_j = S_j \tag{30}$$

So that the j'th component of $\vec{S}$ is simply equal to $\langle U|\boldsymbol{\sigma}_j|U\rangle$. Looking at this from a different standpoint we may express the components of the Stokes vector in terms of the components of $|U\rangle$ in the $(\hat{e}_x, \hat{e}_y)$ representation,

$$S_1 = \langle U|\boldsymbol{\sigma}_1|U\rangle = |u_1|^2 - |u_2|^2 \tag{31}$$

$$S_2 = \langle U|\boldsymbol{\sigma}_2|U\rangle = u_1^*u_2 + u_1u_2^* = |u_{45}|^2 - |u_{-45}|^2 \tag{32}$$

$$S_3 = \langle U|\boldsymbol{\sigma}_3|U\rangle = -iu_1^*u_2 + iu_1u_2^* = |u_R|^2 - |u_L|^2 \tag{33}$$

where we have also expressed S_2 in terms of the components along the linear axes rotated by 45 degrees relative to $\hat{e}_x$ and $\hat{e}_y$ and S_3 in terms of the right circular and left circular components. It is now clear that the circular

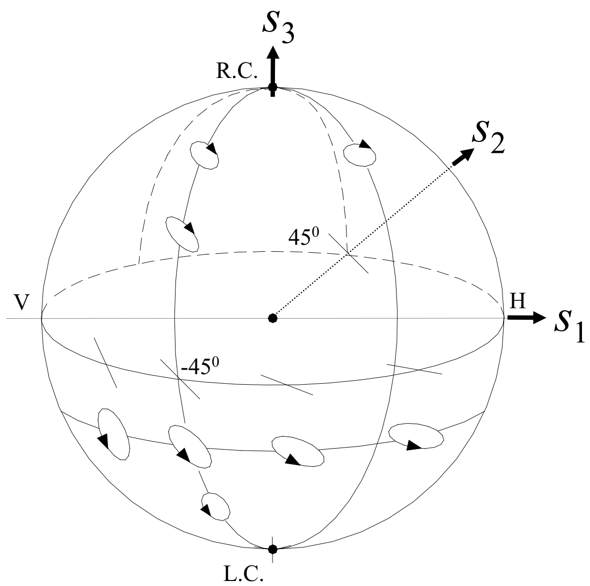

Fig. 1. The relation between the actual polarizations and the position on the Poincaré sphere. The abbreviations H, V, R.C. and L.C. stand for horizontal, vertical, right-circular and left-circular, respectively.

polarizations lie on the poles of the sphere, whereas linear polarizations lie entirely on the equator.

Using the above formalism we may also consider the way that dot products between Jones vectors translate into Stokes space. Clearly we should not expect to preserve the phase of the dot product, but only its absolute value. Consider 2 vectors $|U\rangle$ and $|V\rangle$. Lets look at the absolute square value of their dot product $|\langle U|V\rangle|^2$ that can also be expressed as

$$|\langle U|V\rangle|^2 = \langle U|V\rangle\langle V|U\rangle = \mathrm{Trace}\Big(|V\rangle\langle V||U\rangle\langle U|\Big) \tag{34}$$

Now using (29) and taking advantage of (21), one can easily show that

$$|\langle U|V\rangle|^2 = \frac{|U|^2|V|^2}{2}(1 + \hat{s}_u \cdot \hat{s}_v) \tag{35}$$

where $\hat{s}_u$ and $\hat{s}_v$ are the normalized Stokes vectors that correspond to $|U\rangle$ and $|V\rangle$, respectively.

Transmission through linear optical elements - We have seen that in the Jones picture, transmission through linear unitary optical elements is represented by a 2 by 2 transfer matrix with complex, frequency dependent components. How does it look in the Stokes representation? Before we proceed it is important to note the following: Any 2 by 2 unitary matrix can be expressed in the form $\boldsymbol{T} = \exp(i\boldsymbol{H})$ with $\boldsymbol{H}$ being Hermitian. In particular, this implies that any unitary matrix can be written as

$$\boldsymbol{T} = \exp(i\bar{\varphi})\exp\left(\frac{i}{2}\vec{\beta}\cdot\vec{\boldsymbol{\sigma}}\right) \tag{36}$$

Since we know that the eigenvalues of $\vec{\beta}\cdot\vec{\boldsymbol{\sigma}}$ are $\pm\beta/2$, the value of of the phase difference in equation (16) is $\Delta\varphi = \beta$, and the eigenvectors of $\boldsymbol{T}$ are the SOPs whose normalized Stokes vectors coincide with $\pm\hat{\beta}$. Since the common phase of $\boldsymbol{T}$ has no effect on polarization we will omit it in what follows.

Now the question that we are going to answer is the following. Suppose that $|U\rangle$ is a SOP whose corresponding Stokes vector is $\vec{S}_U$. Suppose that this SOP passes through a linear unitary medium with matrix $\boldsymbol{T} = \exp\left(i\vec{\beta}\cdot\vec{\boldsymbol{\sigma}}/2\right)$ such that at the output we get $|V\rangle = \boldsymbol{T}|U\rangle$. What is the Stokes vector $\vec{S}_v$ that corresponds to $|V\rangle$? The answer will be that $\vec{S}_v$ is obtained by rotating $\vec{S}_u$ about $\vec{\beta}$ in the clockwise direction by an angle equal to β. Lets see how it comes out. The output Stokes vector $\vec{S}_v$ can be explicitly expressed as follows,

$$\vec{S}_v = \langle U|\exp\left(-\frac{i}{2}\vec{\beta}\cdot\vec{\boldsymbol{\sigma}}\right)\vec{\boldsymbol{\sigma}}\exp\left(\frac{i}{2}\vec{\beta}\cdot\vec{\boldsymbol{\sigma}}\right)|U\rangle \tag{37}$$

Expanding the exponent as in (26) and using relations (22),(23) and (25) we have

$$\vec{S}_v = \cos(\beta)(\vec{S}_u - \hat{\beta}(\hat{\beta}\cdot\vec{S}_u)) + \sin(\beta)\hat{\beta}\times\vec{S}_u + \hat{\beta}(\hat{\beta}\cdot\vec{S}_u)$$

$$= \vec{S}_u + \sin(\beta)\hat{\beta} \times \vec{S}_u + (1 - \cos(\beta))\hat{\beta} \times \hat{\beta} \times \vec{S}_u \qquad (38)$$

Looking at this last equation it should not be difficult to recognize that it indeed represents a clockwise rotation of $\vec{S}_u$ about $\hat{\beta}$ by an angle β.

Propagation through a medium with birefringence

So far we have considered transmission through a general linear optical element with a unitary transmission matrix. The transmission matrix could have any arbitrary dependence on frequency and the observations that we made assumed that the optical frequency was kept constant. Now let us introduce the concept of pure birefringence. By this we are referring to a class of optical elements in which we may assume that the two eigen-polarizations are fixed (independent of frequency) and that each can be characterized by a different refractive index such that the index difference $\Delta n = n_1 - n_2$ is independent of frequency. In such elements the accumulation of phase in each one of the eigen-polarizations follows $\varphi_{1,2}(z) = n_{1,2}\omega z/c$ with c being the speed of light. The phase difference between them is $\Delta\varphi(z) = \Delta n\omega z/c$ and it has a linear dependence on both position z and frequency ω. Clearly, the difference in refractive indices can also be viewed as a difference in the propagation delay between the two axes. Thus, injecting a short pulse into a birefringent medium and neglecting dispersion, we expect to find that the pulse splits into two replicas that emerge after different delays from the birefringent medium and whose polarizations coincide with the eigen-polarizations of the birefringent element. From our discussion earlier in this section, it is evident that the matrix representing the action of such a birefringent element will have the form

$$\boldsymbol{T} = \exp\left(\frac{i}{2}\frac{\Delta n\omega z}{c}\hat{n} \cdot \vec{\boldsymbol{\sigma}}\right) \qquad (39)$$

where $\hat{n}$ is a unit Stokes vector that corresponds to the representation of the slow eigen-polarization vector in Stokes space. This also tells us what the effect of pure birefringence looks like in Stokes space when we consider continuous propagation along a birefringent element, or when we consider an element of fixed length, but continuously vary the optical frequency. In the

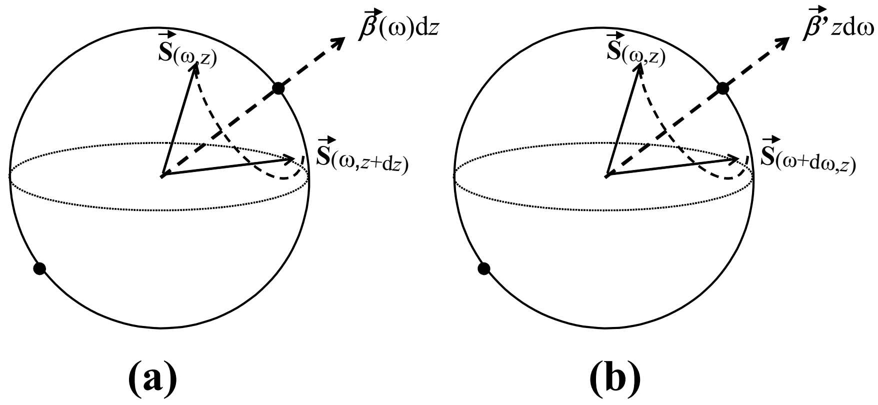

Fig. 2. The effects of changing position (a) along a constant birefringence element and of changing the optical frequency (b) are both reflected through a rotation of the polarization around the birefringence vector in Stokes space. In this example $\vec{\beta}$ is the differential birefringence vector, $\vec{\beta} = (\Delta n\omega/c)\hat{n}$ and $\vec{\beta}' = (\Delta n/c)\hat{n}$ is the frequency derivative of $\vec{\beta}$.

case of long birefringent elements (such as fibers etc.), It is customary to refer to the differential birefringence vector $\vec{\beta} = (\omega\Delta n/c)\hat{n}$, which is the derivative of the overall birefringence vector with respect to the position parameter z.

The effect of Birefringence in optical communications systems - What would happen if a communications channel was to be transmitted through an element with optical birefringence. The result would be quite obvious. The optical signal would emerge as two, orthogonally polarized replicas of itself that are time delayed with respect to each other. Since the optical receiver typically detects the overall power it would sum the two delayed replicas in power and the received signal will often be considerably distorted. Of course, if indeed systems suffered from pure optical birefringence as we are describing here, the distortion could be easily reduced by aligning the input SOP with one of the eigen-polarizations of the system. What makes the problem in actual systems significantly more difficult to deal with is the fact

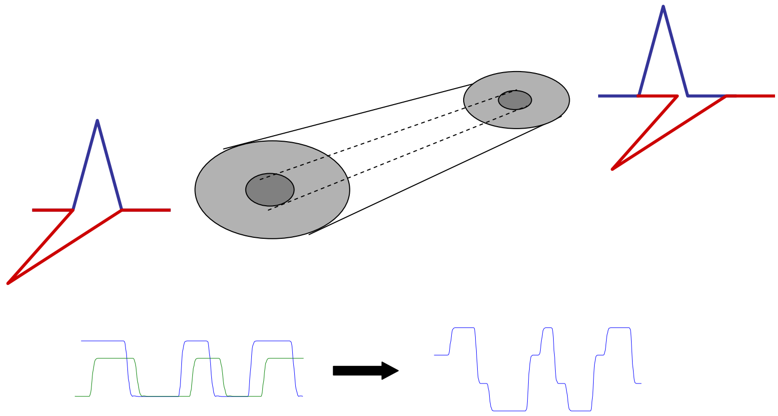

Fig. 3. The effect of pure birefringence on the propagation of a single pulse (top) and an NRZ modulated data-stream.

that the local differential birefringence vector of the fiber, varies randomly with position. This leads to a host of interesting and complicated phenomena that are collectively called polarization mode dispersion or PMD.

3 What is PMD?

Consider a scenario where the local birefringence changes both in its strength (the difference between the refractive indices) and orientation (the eigen-polarizations) along the propagation axis of the optical fiber. To picture what happens in this situation we illustrate in Figure 1b the case in which the birefringence has a stepwise constant behavior. In other words, the fiber consists of many short sections of constant birefringence. In this case, assuming hypothetically that we are launching an optical impulse (see Figure 4a), it splits into two delayed replicas after every individual section such that a very large number of replicas is present at the link output. If instead of an impulse we consider an input pulse of some finite bandwidth, it smears in the process of propagation and becomes distorted in a manner that cannot be easily characterized in the general case. This is in fact the true effect of PMD. It is

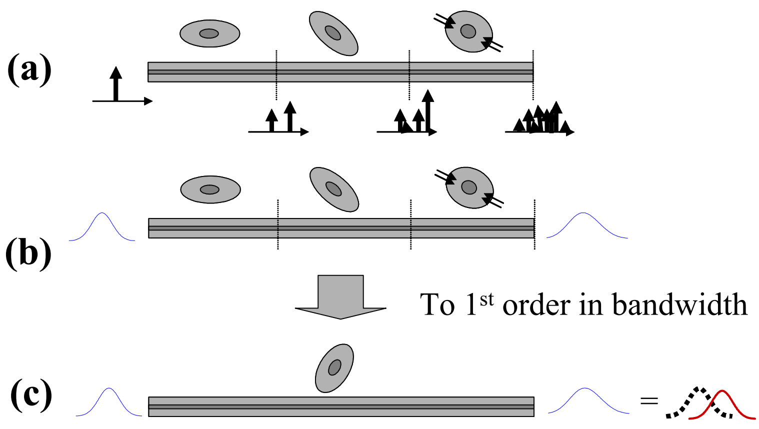

Fig. 4. (a) When the fiber consists of many sections with different birefringence, the transmitted pulse splits many times along the propagation and significant distortions to the transmitted waveform may follow as a result of PMD. (b) When the injected pulse is sufficiently narrow-band. (c) The effect of a fiber with changing birefringence becomes equivalent to first order to the effect of constant birefringence.

interesting that when the bandwidth of the pulse is sufficiently narrow, the distortion that it experiences can be described in terms that are equivalent to the case of constant birefringence. In other words, one could replace the link with a fiber of constant birefringence with some well defined delay and eigen-polarizations that would distort the injected signal in a similar manner. The approximate description of PMD in terms of equivalent birefringence is the so called first order PMD approximation. Generally speaking, it results from the first order expansion of the fiber transmission matrix with respect to frequency, but its justification and range of validity will be discussed in what follows.

We have seen in Figure 2 that in elements with pure birefringence the effect of varying the optical frequency on the output SOP is that of rotation about the birefringent vector at a rate given by the product of the frequency derivative of the birefringence vector and the length (or thickness) of the element $\vec{\beta}'z$. The spectral evolution of the output state of polarization becomes more complicated when the birefringence vector changes along the propagation axis of the fiber. Once again we consider for simplicity the step-wise constant evolution of the birefringence vector which corresponds to the concatenation of many constant birefringence sections with different birefringence vectors. This situation is illustrated in Figure 5a. Figures 5b and 5c describe the evolution of the polarization state at two different optical frequencies when the input state of polarization is kept constant. In the process of passing through each one of the constant birefringence sections, the state of polarization rotates around the corresponding birefringence vectors and draws an arch on the Poincaré sphere. The lengths of the arches are different at the two different frequencies and as a result, the dependence of the output state of polarization on frequency is rather complex. When the number of birefringent sections becomes large the output state of polarization draws a seemingly random curve on the surface of the Poincaré sphere in response to a continuous change in the optical frequency, as is illustrated in Figure 6. Notice, that locally, in the immediate vicinity of any optical frequency ω_0

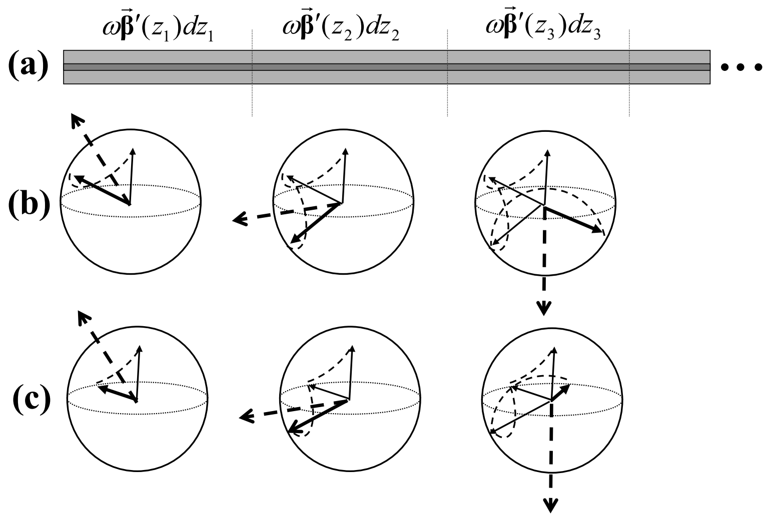

Fig. 5. (a) is an example of a case where the birefringence changes with position along the link. (b) and (c) show what happens to the polarization Stokes vector of the signal when passing through the individual sections. The optical frequency in (c) is slightly lower than in (b) such that all the arches are proportionally shorter. The figure shows that there is no simple relation between the output polarization states that correspond to different frequencies in a complex fiber.

the evolution of the curve can be described as precession[3] about some axis $\vec{\tau}(\omega_0)$. It is customary to set the length of this vector $\vec{\tau}(\omega_0)$ to be equal to the rotation angle of the output state of polarization per unit of frequency in the vicinity of ω_0. When defined in this way $\vec{\tau}(\omega)$ is called the PMD vector, its orientation $\vec{\hat{\tau}}(\omega)$ defines the so called principal states of polarization (PSP)[4] and its magnitude $\tau(\omega_0) = |\vec{\tau}(\omega)|$ is the differential group delay (DGD). The physical meaning of $\vec{\tau}(\omega)$ can be understood through analogy with the bire-

[3] There may be a difficulty with the rigorous meaning of instantaneous axis of rotation in the general case. The meaning here is for illustrative purposes mainly.

[4] More accurately, the principal states are the two polarizations states that are parallel and antiparallel to $\vec{\hat{\tau}}$. It is customary to define the orientation of $\vec{\hat{\tau}}$ such

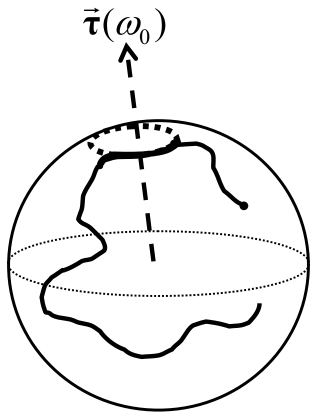

Fig. 6. An illustration of a trajectory that the polarization state at a given position z along the link draws on the Poincaré sphere. Locally the variation of the state of polarization with frequency can be described as rotation around the PMD vector $\vec{\tau}$.

fringence vector. Specifically, if we consider a narrowband optical signal such that within its bandwidth the frequency dependence of $\vec{\tau}$ can be neglected. Then the frequency components of that signal lie on an arch that is part of a circle surrounding the vector $\vec{\tau}(\omega_0)$ with ω_0 being the central optical frequency. In this case the dependence on frequency is identical to what would result from propagation in a constant birefringence fiber whose birefringence vector satisfies $\vec{\beta}'z = \vec{\tau}(\omega_0)$. It is therefore clear by analogy that the output pulse will consist of two orthogonally polarized replicas of the original pulse that are delayed relative to each other. The polarization states of these two replicas coincide with the PSPs and the delay between them is the DGD of the fiber. With this we justified the first order PMD approximation that we initially introduced earlier. We have also seen that this approximation remains valid as long as the PMD vector can be considered constant within the bandwidth of the transmitted signal. This statement will be further quanti-

that the rotation of $\vec{S}$ around it follows the right screw convention. In this case $\hat{\vec{\tau}}$ coincides with the slow PSP.

fied later. To account for the spectral dependence of the PMD vector it is customary to represent it in the form of a Taylor expansion

$$\vec{\tau}(\omega, z) = \vec{\tau}(\omega_0, z) + (\omega - \omega_0)\vec{\tau}'(\omega_0, z) + \frac{1}{2}(\omega - \omega_0)^2\vec{\tau}''(\omega_0, z) + \dots , \quad (40)$$

where the primes over $\vec{\tau}$ denote derivatives with respect to frequency. The term $\vec{\tau}(\omega_0, z)$ represents the first order PMD approximation since it neglects the spectral variation of PMD altogether. The frequency derivatives of $\vec{\tau}$ represent the so-called high orders of PMD, where it has become customary to associate the n'th frequency derivative of the PMD vector with the $n + 1$ PMD order. Other, alternative, definitions of high orders of PMD have also been proposed recently[5].

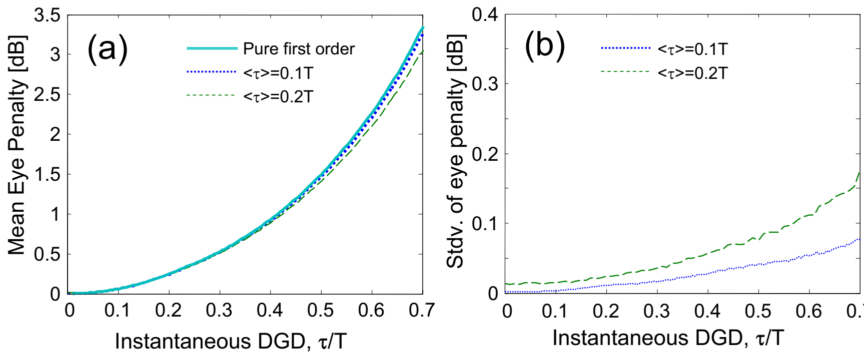

Fig. 7. The mean (a) and the standard deviation (b) of the eye penalty of a typical RZ modulated system (50% duty cycle) as a function of the instantaneous first order PMD normalized to the symbol duration. The average Differential group delay is indicated in the legend.

In order to assess the value of the first order PMD approximation in communications systems one can refer to Figure[6] 7 showing the result of a large

[5] see for example Eyal et. al. Electron. Lett. **19**, 1658 (1999).

[6] A discussion of this figure and its detailed description can be found in Shtaif et. al. IEEE Photon. Technol. Lett., p.1369, 2003.

number of simulations in which the link is modelled as a concatenation of 100 statistically independent birefringent sections with an identical distribution of the birefringence vector in each section. The modulation format is RZ with $\sim 50\%$ duty cycle. Figure 7a shows the eye penalty evaluated in a simulation of a standard fiber-optic system with PMD, as a function of the instantaneous DGD. The dashed and dotted curves (blue and green) in this figure correspond to the case where the mean DGD of the link is 0.1 and 0.2 times the symbol duration. The cyan curve represents what would be the penalty in the case of pure (position independent) birefringence where high-order PMD is absent. The difference between the 3 curves is caused only by whatever concerns the high-order PMD effects. It is obvious that the contribution of high order effects in these examples is very small. This can also be deduced from Figure 7b showing that the standard deviation of the eye penalty in the various cases (which are caused exclusively by the existence of high orders) is also small relative to the first order penalty. From all of the above it is evident that knowledge of the PMD vector allows a very simple prediction of the expected penalty with a fairly good accuracy based on the first-order PMD approximation. Nevertheless, to put things into a correct perspective, it must be mentioned that in various cases when rare error-events need to be considered, or when there is high average DGD (larger than 20% of the symbol duration) the contribution of high-orders of PMD is important and needs to be taken into account.

The PMD equations

The above description can be mathematically formulated in the following way,

$$\vec{S}(\omega, z + \mathrm{d}z) = \mathsf{R}(\vec{\beta}(\omega, z)\mathrm{d}z)\vec{S}(\omega, z) \tag{41}$$

$$\vec{S}(\omega + \mathrm{d}\omega, z) = \mathsf{R}(\vec{\tau}(\omega, z)\mathrm{d}\omega)\vec{S}(\omega, z) \tag{42}$$

where $\mathsf{R}(\vec{V})$ is a rotation operator that rotates the vector that it acts upon by an angle $|\vec{V}|$ around the axis $\hat{\vec{V}}$. This operator can also be represented in an exponential operator form as $\mathsf{R}(\vec{V}) = \exp(\vec{V}\times)$, with $(\vec{V}\times)$ representing the operation of a cross product. When the rotation operators are expanded to first order with respect to their argument ($\mathsf{R}(\vec{\beta}\mathrm{d}z) \simeq 1 + \mathrm{d}z\vec{\beta}\times$ and $\mathsf{R}(\vec{\tau}\mathrm{d}\omega) \simeq 1 + \mathrm{d}\omega\vec{\tau}\times$) the more familiar form of expressions (41) and (42) follows

$$\frac{\partial \vec{S}}{\partial z} = \vec{\beta} \times \vec{S} \tag{43}$$

$$\frac{\partial \vec{S}}{\partial \omega} = \vec{\tau} \times \vec{S}. \tag{44}$$

Furthermore, by combining equations (41) and (42) and using the algebra of rotations, it can also be shown (see appendix 3) that the evolution of $\vec{\tau}(\omega, z)$ is described by

$$\vec{\tau}(\omega, z + \mathrm{d}z) = \mathsf{R}(\vec{\beta}(\omega, z)\mathrm{d}z)\vec{\tau}(\omega, z) + \vec{\beta}'(\omega, z)\mathrm{d}z, \tag{45}$$

where we recall that $\vec{\beta}'$ is the delay vector which is equal to the frequency derivative of $\vec{\beta}(\omega, z)$. We should stress that while the vector $\vec{\tau}(\omega, z)$ in equation (45) reflects the PMD that is accumulated between the beginning of the link and point z, the vectors $\vec{\beta}(\omega, z)\mathrm{d}z$ and $\vec{\beta}'(\omega, z)\mathrm{d}z$ represent the birefringence and delay, respectively, of the section of fiber that starts at z and ends at $z + \mathrm{d}z$. This clarification becomes important when the birefringence vector is characterized by abrupt changes along the link as in the case where the fiber is constructed from many sections of constant birefringence. In the continuous case, using equations (43) and (44) one can obtain (make sure you see how it is done) the following evolution equation for the PMD vector

$$\frac{\partial \vec{\tau}}{\partial z} = \vec{\beta} \times \vec{\tau} + \beta'. \tag{46}$$

Taking an additional view of the PMD problem we may write the following equation for the transfer matrix of PMD.

$$\frac{\partial \boldsymbol{T}}{\partial \omega} = i\left(\tau_0 \mathbf{1} + \frac{1}{2}\vec{\tau}(\omega) \cdot \vec{\boldsymbol{\sigma}}\right)\boldsymbol{T}(\omega) \tag{47}$$

where the hermitian form of the operator in parentheses is necessary to preserve the unitary nature of the Jones matrix $\boldsymbol{T}$ as the frequency varies. Notice that τ_0 is nothing but a fixed, polarization independent delay term. It may in principle depend on ω in which case it also reflects the effect of chromatic dispersion. Let me now show you that the PMD vector $\vec{\tau}$ appearing in (47) is the same one appearing in (46). Assume that we start with a polarized input SOP $|U\rangle$ and end up with $|V\rangle = \boldsymbol{T}(\omega)|U\rangle$. The output Stokes vector is $\vec{S}_v = \langle V|\vec{\boldsymbol{\sigma}}|V\rangle = \langle U|\boldsymbol{T}^\dagger \vec{\boldsymbol{\sigma}} \boldsymbol{T}|U\rangle$. The frequency derivative of $\vec{S}_v$ is

$$\begin{aligned}
\frac{\partial \vec{S}}{\partial \omega} &= \langle U|(\boldsymbol{T}^\dagger)' \vec{\boldsymbol{\sigma}} \boldsymbol{T}|U\rangle + \langle U|\boldsymbol{T}^\dagger \vec{\boldsymbol{\sigma}} \boldsymbol{T}'|U\rangle \\
&= i\langle V| \left(\tau_0 \mathbf{1} + \frac{1}{2}\vec{\tau}\cdot\vec{\boldsymbol{\sigma}}\right) \vec{\sigma}|V\rangle - i\langle V|\vec{\sigma} \left(\tau_0 \mathbf{1} + \frac{1}{2}\vec{\tau}\cdot\vec{\boldsymbol{\sigma}}\right)|V\rangle \\
&= \langle V|\vec{\tau}\times\vec{\boldsymbol{\sigma}}|V\rangle = \vec{\tau}\times\vec{S}
\end{aligned} \tag{48}$$

where for the last stage we used relations (22) and (23).

Equation (47) gives us a simple way of finding the PMD vector out of the Jones matrix. In analogy with what we have seen before,

$$\tau_j = \mathrm{Trace}\left(-i\frac{\partial \boldsymbol{T}}{\partial \omega}\boldsymbol{T}^\dagger \boldsymbol{\sigma}_j\right) \tag{49}$$

where τ_j is the j'th component of the PMD vector ($j = 1, 2, 3$).