Stokes-space analysis of modal dispersion in fibers with multiple mode transmission

Cristian Antonelli,[1,*] **Antonio Mecozzi,**[1] **Mark Shtaif,**[2] **and Peter J. Winzer**[3]

[1] *Department of Electrical and Information Engineering, University of L'Aquila, 67100 L'Aquila, Italy*
[2] *School of Electrical Engineering, Tel Aviv University, Tel Aviv, 69978 Israel*
[3] *Bell Labs, Alcatel-Lucent, 791 Holmdel-Keyport Rd., Holmdel, New Jersey 07733, USA*
[*]*cristian.antonelli@univaq.it*

Abstract: Modal dispersion (MD) in a multimode fiber may be considered as a generalized form of polarization mode dispersion (PMD) in single mode fibers. Using this analogy, we extend the formalism developed for PMD to characterize MD in fibers with multiple spatial modes. We introduce a MD vector defined in a D-dimensional extended Stokes space whose square length is the sum of the square group delays of the generalized principal states. For strong mode coupling, the MD vector undertakes a D-dimensional isotropic random walk, so that the distribution of its length is a chi distribution with D degrees of freedom. We also characterize the largest differential group delay, that is the difference between the delays of the fastest and the slowest principal states, and show that it too is very well approximated by a chi distribution, although in general with a smaller number of degrees of freedom. Finally, we study the spectral properties of MD in terms of the frequency autocorrelation functions of the MD vector, of the square modulus of the MD vector, and of the largest differential group delay. The analytical results are supported by extensive numerical simulations.

© 2012 Optical Society of America

OCIS codes: (060.2330) Fiber optics communications; (060.4510) Optical communications; (060.4230) Multiplexing.

References and links

1. A. R. Chraplyvy, "The coming capacity crunch," European Conference on Optical Communication 2009 (ECOC09), plenary talk (2009).
2. R. W. Tkach, "Scaling optical communications for the next decade and beyond," Bell Labs Tech. J. **14**, 3–9 (2010).
3. Peter J. Winzer and Gerard J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," Opt. Express **19**, 16680–16696 (2011).
4. S. Chandrasekhar, A. H. Gnauck, X. Liu, P. J. Winzer, Y. Pan, E. C. Burrows, B. Zhu, T. F. Taunay, M. Fishteyn, M. F. Yan, J. M. Fini, E. M. Monberg, and F. V. Dimarcello, "WDM/SDM transmission of 10×128-Gb/s PDM-QPSK over 2688-km 7-core fiber with a per-fiber net aggregate spectral-efficiency distance product of 40,320 km$\times$b/s/Hz," Proceedings of European Conference on Optical Communication 2011 (ECOC11), Paper Th.13.C4 (2011).
5. S. Randel, R. Ryf, A. Sierra, P.J. Winzer, A.H. Gnauck, C.A. Bolle, R-J. Essiambre, D.W. Peckham, A. McCurdy, and R. Lingle "6×56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6×6 MIMO equalization," Opt. Express **19**, 16697–16707 (2011).

6. E. Ip, N. Bai, Y.K. Huang, E. Mateo, F. Yaman, S. Bickham, H.Y. Tam, C. Lu, M.J. Li, S. Ten, A.P. Tao Lau, V. Tse, G.D. Peng, C. Montero, X. Prieto, and G. Li, "88 × 3 × 112-Gb/s WDM transmission over 50 km of three-mode fiber with inline few-mode fiber amplifier," Proceedings of European Conference on Optical Communication 2011 (ECOC11), Paper Th.13.C2 (2011).
7. S. Fan and J. M. Kahn, "Principal modes in multi-mode waveguides", Opt. Lett. **30**, 135–137 (2005).
8. K-P. Ho and J. M Kahn, "Statistics of group delays in multi-mode fibers with strong mode coupling," J. Lightwave Technol. **29**, 3119–3128 (2011).
9. K-P. Ho and J .M Kahn, "Mode-dependent loss and gain: statistics and effect on mode-division multiplexing," Opt. Express **19**, 16612–16635 (2011).
10. J. P. Gordon and H. Kogelnik, "PMD fundamentals: polarization mode dispersion in optical fibers," Proc. Natl. Acad. Sci. USA **97**, 4541–4550 (2000).
11. N. Gisin, "Statistics of polarization dependent losses," Opt. Commun. **114**, 399–405 (1995).
12. D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, "Measurement of qubits," Phys. Rev. A **64**, 052312 (2001).
13. M. Nazarathy and E. Simony, "Generalized Stokes parameters shift keying approach to multichip differential phase encoded optical modulation formats," Opt. Lett. **31**, 435–437 (2006).
14. M. Salsi, C. Koebele, D. Sperti, P. Tran, H. Mardoyan, P. Brindel, S. Bigo, A. Boutin, F. Verluise, P. Sillard, M. Astruc, L. Provost, and G. Charlet, "Mode division multiplexing of 2 × 100Gb/s channels using an LCOS based spatial modulator," J. Lightwave Technol. **30**, 618–623 (2012).
15. S. Randel, M. Magarini, R. Ryf, R-J. Essiambre, P. J. Winzer, T. Hayashi, T. Taru, and T. Sasaki "MIMO-based signal processing of spatially multiplexed 112-Gb/s PDM-QPSK signals using strongly-coupled 3-core fiber," Proceedings of European Conference on Optical Communication 2011 (ECOC11), Paper Tu.5.B1 (2011).
16. D. Cassioli, M. Z. Win, and A. F. Molisch, "The ultra-wide bandwidth indoor channel: from statistical model to simulations," IEEE J. Sel. Areas Commun. **20**, 1247–1257 (2002).
17. In the literature the shape parameter K_N is typically referred to as the number of degrees of freedom of the chi distribution. We refrain from this terminology so as to avoid confusion with the number of degrees of freedom D_{eff} representing the dimension of the manyfold spanned by the MD vector in the generalized Stokes space .
18. M. Karlsson and J. Brentel, "The autocorrelation function of the polarization-mode dispersion vector," Opt. Lett. **24**, 939–941 (1999).
19. M. Shtaif and A. Mecozzi, "Study of the frequency autocorrelation of the differential group delay in fibers with polarization mode dispersion," Opt. Lett. **25**, 707–709 (2000).
20. R. Ryf, S. Randel, A. H. Gnauck, C. Bolle, A. Sierra, S. Mumtaz, M. Esmaeelpour, E. C. Burrows, R-J. Essiambre, P. J. Winzer, D. W. Peckham, A. H. McCurdy, and R. Lingle, "Mode-division multiplexing over 96km of few-mode fiber using coherent 6 ×6 MIMO processing," J. Lightwave Technol. **30**, 521–531 (2012).
21. M. Gell-Mann, "Symmetries of baryons and mesons," Phys. Rev. **125**, 1067–1084 (1962).
22. A. Galtarossa, L. Palmieri, M. Schiano, and T. Tambosso, "Measurement of birefringence correlation length in long, single-mode fibers," Opt. Lett. **26**, 962–964 (2001).
23. F. Curti, B. Daino, G. De Marchis, and F. Matera, "Statistical treatment of the evolution of the principal states of polarization in single-mode fibers," J. Lightwave Technol. **8**, 1162–1166 (1990).
24. The more common terminology for the vector operation defined in Eq. (28) is Lie bracket. We prefer the term generalized cross-product for emphasizing the analogy with the case of PMD.

1. Introduction

With the fiber-optic communications industry facing an imminent capacity crunch [1, 2], the search for new, scalable technologies has become unavoidable. One of the most promising approaches for preventing the predicted crisis is the use of spatially multiplexed transmission [3] in a combination of multi-core (MCF) [4] and multi-mode (MMF) fibers [5, 6]. Indeed, a notable fraction of regular and post-deadline papers presented in the last two largest conferences on optical communications (the Optical Fiber Communications conference (OFC), Los Angeles 2011, and the European Conference on Optical Communications (ECOC), Geneva 2011) were devoted to spatially multiplexed optical transmission.

Conceptually, provided that all fiber modes are selectively addressed at transmitter and receiver, spatially multiplexed fiber-optic transmission is a direct extension of polarization multiplexed transmission in single mode fibers that is being used today. To describe a fiber with N spatial modes, the 2-dimensional polarization vector needs to be replaced by a $2N$ dimensional vector whose elements represent the excitation of the various modes (with the factor of 2 ac-

counting for the fact that each mode supports two polarizations). Correspondingly, the 2×2 transfer matrix that describes linear transmission in single mode fibers needs to be replaced by a matrix whose dimension is $2N \times 2N$. This situation calls for the extension of the concepts that are at the heart of polarization studies. Thus, rather than referring to the phenomena of polarization mode dispersion (PMD) and polarization dependent loss (PDL), which describe the relevant consequence of polarization coupling, one now has to deal with the more general concepts of modal dispersion (MD) and mode dependent loss (MDL). Relevant work on this topic has already been reported, where the concept of principal states of polarizations was generalized to the multi-mode case [7] and where the statistics of the corresponding group delays was derived [8]. Important observations on the dynamics of MDL were also reported in [9]. Nonetheless, the understanding of the physical nature of these phenomena can greatly benefit from the generalization of the analytical tools, which were invaluable in polarization studies.

Polarization vectors, which have two complex-number components, reside in a so-called Jones space. This Jones space is isomorphic to a so-called Stokes space which consists of 3 dimensional real-valued column vectors [10]. Thus, every state of polarization can be uniquely represented either as a Jones vector, or as a vector in Stokes space. In addition, unitary operations in Jones space can be represented by a Stokes vector, whose direction is the axis of rotation and whose modulus equals the rotation angle. For this reason, the effect of birefringence (which is a unitary operation in Jones space) is customarily represented by a 3-dimensional vector $\vec{\beta}$ in Stokes space, known as the *birefringence vector*. Specifying the birefringence of a fiber therefore comes down to specifying the birefringence vector $\vec{\beta}(z)$ that represents it at every point z along the fiber. Similarly, the effect of PMD is customarily represented by a vector $\vec{\tau}$ in Stokes space, known as the PMD vector, capturing birefringence-induced waveform distortion. The Stokes representations of the slowest and fastest states of polarization (known as the principal states of polarization) coincide with $\pm \hat{\tau}$ (where $\hat{\tau} = \vec{\tau}/\tau$ is a unit vector), and the differential group delay (DGD) between the slowest and fastest polarization components is equal to $\tau = |\vec{\tau}|$. Even non-unitary phenomena such as PDL are typically represented in terms of vectors in Stokes space [11]. It is thus natural for the analysis of spatially multiplexed transmission to seek an extension of the above concepts to the multi-mode case.

In this paper, we introduce an extended $D = 4N^2 - 1$ dimensional Stokes space, that allows the representation of multi-mode $2N$-dimensional state-vectors (which are the extension of the 2-dimensional Jones vectors in the case of polarization). While generalized Stokes representations have been considered previously in other contexts [12, 13], the properties of the generalized Stokes space and the relation to unitary dynamics (which is of particular importance in our case) have not been fully formulated previously, to the best of our knowledge. We show that while the extended Stokes-space lacks some of the properties of the usual Stokes-space describing polarization, many properties can be generalized in a way that sheds light on the physics of multi-mode propagation. In particular, focusing on the case where MDL is negligible, we show that the local birefringence vector $\vec{\beta}$ and the PMD vector $\vec{\tau}$ can be extended to the D-dimensional space so as to capture the effects of local mode coupling and MD. As in the case of PMD, the MD vector $\vec{\tau}$ determines the temporal spread of propagating pulses, although the relation between its magnitude and the differential mode delays is somewhat less direct in the multi-mode case, as we shall see. In order to demonstrate the use of the proposed formalism we consider the case of quasi-degenerate modes, namely transmission over multiple spatial modes with varying degrees of coupling, but with small deterministic difference between the various group delays, such that the delay spread observed at the output is always dominated by the random coupling caused by perturbations. In addition, our theory applies to the case of arbitrary deterministic group delay difference, as long as the coupling between the modes is strong. The statistics of mode dispersion in this scenario are identical to that of strongly coupled degenerate

modes. This framework includes real-world scenarios in which groups of modes propagating at different group velocities are processed separately [14]. For example, this corresponds to the case in which the four degenerate LP_{11} modes of a step-index fiber are processed separately from its two LP_{01} modes. Larger groups of degenerate modes can be encountered in situations where multiple MMFs are combined in a MCF structure. Our analysis does not rely on a specific choice of the set of modes that are used to describe signal propagation in the optical fiber and it is valid as long as the quasi-degeneracy conditions are satisfied. For example, in the case of a single-core multi-mode fiber one could choose to work with either the exact (hybrid) fiber modes, or with the linearly polarized mode approximation, with no consequence to the validity of the results.

The source of MD within a group of degenerate modes is random fiber imperfection and environmental perturbation that remove the degeneracy and cause group-delay spread. While MD can in principle be compensated for using digital signal processing (DSP) [15], the requirements from the DSP hardware and algorithm, as well as the overall system performance strongly depend on the statistical properties of MD, which we extract in this paper. In what follows we will use the acronym MMF in order to refer collectively to all structures in which light can propagate in multiple spatial modes, whether by means of a single multi-mode waveguide, or through multiple cores.

2. Main results

Prior to delving into the analytical details of the proposed model, we devote the present section to a summary of the main results.

2.1. Generalized Stokes space and modal dispersion vector

An important outcome of this work is the introduction of a generalized D-dimensional Stokes space (with $D = 4N^2 - 1$) and the definition of a D-dimensional MD vector $\vec{\tau}$, as derived in Section 3. Knowledge of $\vec{\tau}$ allows the extraction of the principal modes of the system as well as their various group delays, although the relations between those quantities and $\vec{\tau}$ are less transparent than they are in the case of PMD. As shown in Section 3.2, the quantity τ^2, which is the square modulus of $\vec{\tau}$, is proportional to the sum of the squares of the $2N$ individual group delays t_i,

$$\tau^2 = 2N \sum_{i=1}^{2N} t_i^2. \tag{1}$$

The individual group delays are defined with the mode averaged delay set to 0, $\sum_j t_j = 0$. The MD vector $\vec{\tau}$ evolves as a Gaussian vector, such that its modulus is chi distributed (its square modulus obeys the chi-square distribution). If no spatial mode coupling exists (while still allowing for polarization coupling within each spatial mode), only $3N$ of the components of $\vec{\tau}$ are different from zero, and the distribution of the modulus τ is chi distributed with $3N$ degrees of freedom. In the other extreme, i.e. in the presence of significant random coupling among all N spatial modes (as well as among the 2 polarizations within each spatial mode), the number of degrees of freedom is $4N^2 - 1$ implying that the orientation of $\vec{\tau}$ is homogenously distributed in the entire generalized Stokes space, and the chi distribution has $4N^2 - 1$ degrees of freedom. Note that the chi distribution is a special case of the Nakagami distribution, which is famous in the context of wireless communications [16]. Interestingly, in the presence of partial coupling, the distribution of τ can be well approximated as a chi distribution with an effective number of degrees of freedom D_{eff}, ranging between $3N$ and $4N^2 - 1$. This transition is illustrated in Fig. 1, where we plot the probability density function (PDF) of τ for different coupling regimes for the case $N = 2$. In (a) the spatial modes are uncoupled ($D_{\text{eff}} = 3N$) and in
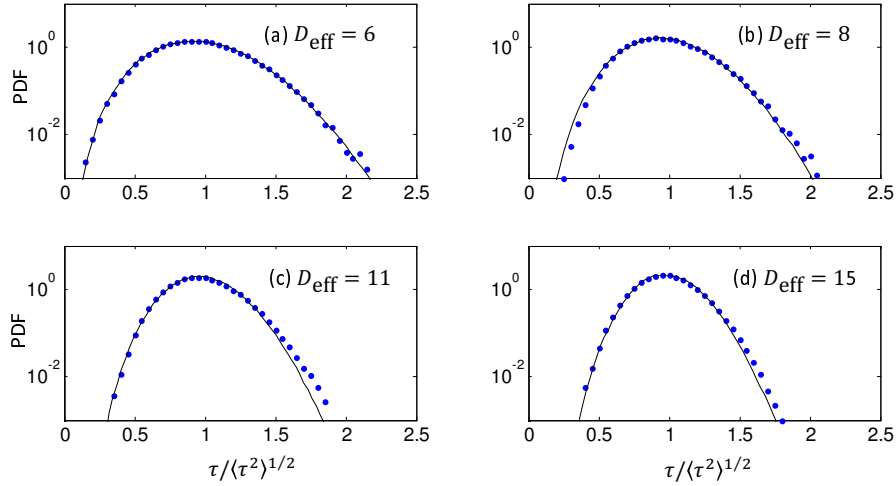
Fig. 1. Probability density function of τ for $N = 2$, normalized to its root-mean square value. The coupling magnitude increases from no coupling in (a) to full coupling in (d). A similar behavior has been verified for a number of larger values of N.

(d) there is full coupling ($D_{\text{eff}} = 4N^2 - 1$); plots (b) and (c) represent different degrees of partial coupling. The dots are the results of Monte Carlo simulations and the solid curves correspond to chi probability density functions with D_{eff} degrees of freedom. The details of the computation are discussed in Section 4.

2.2. Maximum modal delay spread

A quantity of particular relevance in the analyzed scenario is the largest *differential* group delay, defined as the difference between the largest and the smallest group delays, $T = \max_k\{t_k\} - \min_k\{t_k\}$, as discussed in Section 3.2. This parameter defines the maximum spread of the propagating waveforms and is therefore related to the required complexity of the digital signal processing hardware (e.g. the number of taps of a finite impulse response equalizer). While the derivation of the distribution of T in the general case seems to be a very difficult task [8, 9], the chi distribution turns out to provide an excellent approximation, as illustrated in Fig. 2 for the case of full coupling. The parameters $\langle T^2 \rangle$ and K_N characterizing the distribution were obtained numerically and are plotted in the top panel of Fig. 2 as a function of the number of spatial modes N. The term $\langle T^2 \rangle$ is the mean square value of T, whereas K_N is what we refer to as the *shape parameter* of the chi-distribution used to fit the PDF of T [17]. The dependence of $\langle T^2 \rangle$ and K_N on N is very well approximated by the following heuristic expressions

$$K_N \simeq 3 + \left\lceil 10.39 \times (N-1)^{1.36} \right\rceil \tag{2}$$

$$\langle T^2 \rangle \simeq \frac{(N-1)^2 + 24.7 \times (N-1) + 16.14}{0.2532 \times (N-1)^2 + 7.401 \times (N-1) + 16.14} \langle \tau^2 \rangle / N^2, \tag{3}$$

where $\lceil x \rceil$ denotes the smallest integer greater than x. The approximate curves are not displayed in Fig. 2, as they are practically indistinguishable from the actual curves in the plotted scale. The curves in the bottom panels illustrate the accuracy of this approximation on a logarithmic scale for three values of N. The plots show that, as the number of spatial modes increases, the chi distribution becomes more and more symmetric around the root-mean square value of T,
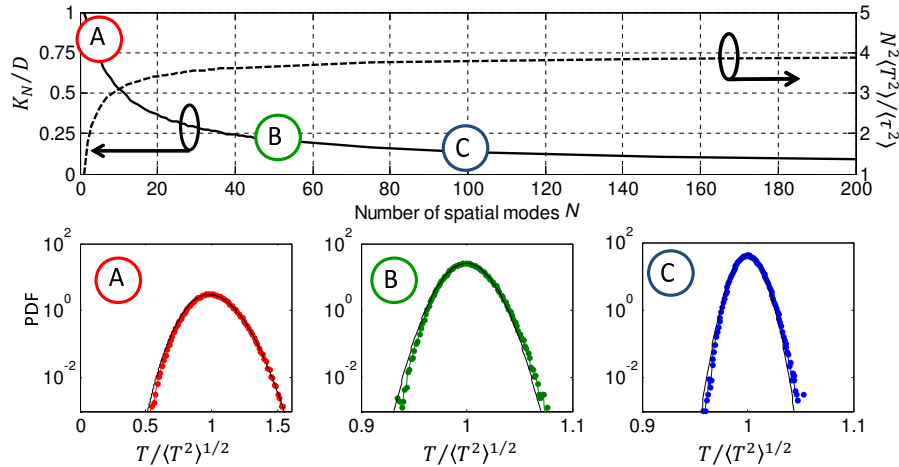
Fig. 2. Top panel. The solid line represents the shape parameter K_N of the chi PDF that provides the best approximation of the distribution of the largest DGD T as a function of N; the dashed line shows the mean square value of T normalized to $\langle \tau^2 \rangle / N^2$. Bottom panels. Probability density function of T for (A) $N = 3$, (B) $N = 50$ and (C) $N = 100$. Dots are the results of Monte-Carlo simulations, solid lines are the plot of chi PDFs with parameters taken from the top panel.

asymptotically approaching the Gaussian distribution. The analytic expression provided in [8] for the case of $2N = 3$ (which unfortunately does not correspond to a practical multi-mode scenario) also agrees very closely with the chi approximation, the differences between the two curves being practically unnoticeable.

2.3. Autocorrelation functions

The spectral behavior of MD is conveniently quantified in terms of the autocorrelation functions of the MD-vector and of its modulus. The width of these functions is a measure of the bandwidth across which the distortion affecting different spectral components of the transmitted optical field are statistically correlated, and it is also defined as the bandwidth across which the first-order approximation [10] applies to the interpretation of MD effects. In the regime of strong coupling we were able to derive analytic expressions for the frequency autocorrelation function of the MD vector $\vec{\tau}$,

$$R_{\vec{\tau}}(\omega) = \langle \vec{\tau}(\Omega + \omega) \cdot \vec{\tau}(\Omega) \rangle = \frac{D}{\omega^2} \left[1 - \exp\left(-\frac{\omega^2 \langle \tau^2 \rangle}{D} \right) \right], \tag{4}$$

as well as that of its square modulus τ^2,

$$R_{\tau^2}(\omega) = \langle \tau^2(\Omega + \omega) \tau^2(\Omega) \rangle = \langle \tau^2 \rangle^2 + \frac{4 \langle \tau^2 \rangle}{\omega^2} - \frac{4D}{\omega^4} \left[1 - \exp\left(-\frac{\omega^2 \langle \tau^2 \rangle}{D} \right) \right]. \tag{5}$$

These expressions generalize the expressions obtained in the analysis of PMD [18, 19] to which they reduce in the case of $N = 1$, that is $D = 3$. The ACF of $\vec{\tau}$, Eq. (4), normalized to its peak value, is plotted versus $\omega(\langle \tau^2 \rangle / D)^{1/2}$ by solid lines in the top panels of Fig. 3 for $N = 2$, $N = 3$ and $N = 4$. The stars are the results of Monte Carlo simulations that we performed to test the analytical result. The corresponding ACF of τ^2, Eq. (5), after subtraction of its asymptotic value for $\omega \to \infty$ and normalization to its peak value, is plotted by solid lines in the bottom panels of
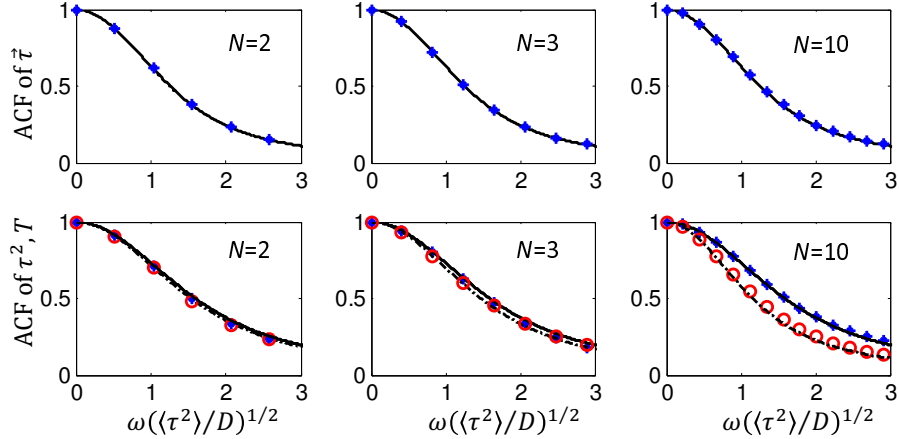
Fig. 3. Top panels: normalized autocorrelation function of $\vec{\tau}$ according to theory (solid curve) and simulations (stars). Bottom panels: the normalized autocorrelation function of τ^2 according to theory (solid curve) and simulations (stars) as well as the autocorrelation function of T (circles). Dot-dashed curves represent the re-scaled ACF of τ^2 Eq. (5), modified by replacing D with K_N as from Fig. 2.

Fig. 3, where the stars are again the results of Monte Carlo simulations. We numerically verified that the autocorrelation function of τ, $R_\tau(\omega) = \langle \tau(\Omega + \omega)\tau(\Omega) \rangle$, whose analytic expression is not available, is approximated very well by Eq. (5), provided that their asymptotic values are removed and that they are normalized to their peak values. More remarkably, we verified that Eq. (5), re-scaled and modified by replacing D with K_N, gives an excellent approximation of the autocorrelation function of the largest *differential* group delay. The ACF of T is plotted in the bottom panels of Fig. 3 by circles and the modified Eq. (5) by dot-dashed lines: although the modification of the re-scaled ACF of τ^2 is practically unnoticeable for small values of N, an excellent match between simulations and analytical expressions characterizes all cases. The bandwidth of the largest *differential* group delay B_T can thus be very well approximated as the 3dB width of the modified ACF (5) $B_T \simeq 3.2\sqrt{K_N/\langle\tau^2\rangle}$, which, for a moderate, yet practically important, number of spatial modes ($\sim N < 5$), is only negligibly different from $B_T \simeq 3.2\sqrt{D/\langle\tau^2\rangle}$.

Analytical expressions corresponding to Eq. (4) and (5) in the regime of partial coupling are not available. However, we found through numerical simulations that the intriguing result illustrated in Fig. 1 (i.e., the good match to a chi distribution independent of D_{eff}) extends somewhat to the MD correlation properties. Indeed, the shape of the autocorrelation function of $\vec{\tau}$ and τ^2 is not affected by the coupling magnitude, whereas its bandwidth increases as the coupling magnitude increases and scales, yet only approximately, with $\sqrt{D_{\text{eff}}}$.

2.4. Non-degenerate modes

Figures 1(d), 2 and 3 were plotted for the case of N degenerate fully coupled modes with no MDL. This situation differs somewhat from reality, where there are groups of degenerate modes that are characterized by deterministically different wave-numbers and hence they couple to each other significantly less as a result of perturbations. Using the characteristic example of the coupling between the LP_{01} and the LP_{11} groups of modes in step-index fibers [20], we show that the results presented above remain practically unchanged. To illustrate this point, we show in Fig. 4(a) the PDF of the length of the $D_{LP_{11}} = 15$ dimensional MD vector $\vec{\tau}_{LP_{11}}$ that
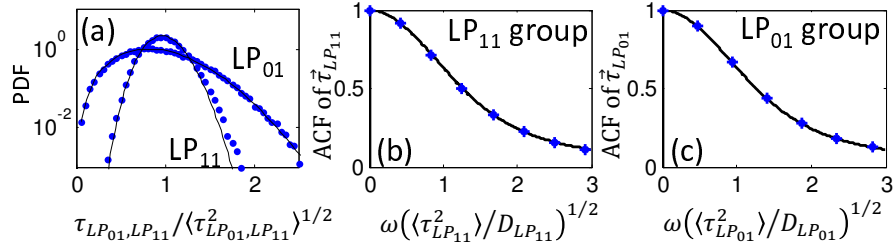
Fig. 4. (a) Probability density function of the length of the MD vector $\vec{\tau}_{LP_{11}}$ that represents the four LP_{11} modes (subspace dimension $D_{LP_{11}} = 15$) and of the length of the MD vector $\vec{\tau}_{LP_{01}}$ that represents the two LP_{01} modes (subspace dimension $D_{LP_{11}} = 3$). The degree of coupling between the two groups of modes is discussed in the text. Figures (b) and (c) show the ACFs of the the two MD vectors. Symbols refer to numerical simulations, solid lines to the theory.

represents the four LP_{11} modes and the PDF of the length of the $D_{LP_{01}} = 3$ dimensional MD vector $\vec{\tau}_{LP_{01}}$ representing the two polarization modes of LP_{01}. In Figs. 4(b) and 4(c) we show the corresponding MD vector ACFs. In order to plausibly emulate the situation in a real fiber [20], we adjusted the magnitude of the difference between the wave-numbers corresponding to the two groups of modes to the point where the overall output power in the LP_{11} modes was about 10dB lower than the output power in the LP_{01} mode, given that only the LP_{01} mode was excited. As is evident from the figures, the numerically obtained distribution of $\tau_{LP_{11}}$ and $\tau_{LP_{01}}$ (symbols) is still very well fitted by chi distributions of $D_{LP_{11}} = 15$ and $D_{LP_{01}} = 3$ degrees of freedom respectively (solid lines). At the same time the numerical ACF (symbols) still matches the theoretical curves (solid lines) given by Eq. (4) very well. The individual MD vectors were extracted by considering the unitary part of the corresponding 4×4 and 2×2 blocks along the main diagonal of the simulated overall (6×6) transfer matrix.

3. Theory

Following [10], we use the notation $|\psi(z, \omega)\rangle$ to represent the state of the field in what we refer to as the generalize Jones space, at angular frequency ω at a given point z along the optical fiber. One should interpret $|\psi(z)\rangle$ as a column vector with $2N$ complex components representing the excitation of the various modes in the fiber. The corresponding row vector, which is the hermitian conjugate of $|\psi(z)\rangle$, is denoted by $\langle\psi(z)| = (|\psi(z)\rangle)^\dagger$. For convenience we will also assume in what follows that the state vector $|\psi(z)\rangle$ is normalized, such that $\langle\psi(z)|\psi(z)\rangle = 1$. Propagation between the positions z_0 and z along the fiber is represented by a linear operator $\mathbf{U}(z, z_0)$ so that

$$|\psi(z)\rangle = \mathbf{U}(z, z_0)|\psi(z_0)\rangle, \tag{6}$$

where the dependence of $\mathbf{U}(z, z_0)$ and $|\psi(z)\rangle$ on ω was suppressed for brevity of notation. When considering the process of propagation in the frequency domain, $\mathbf{U}(z, z_0)$ can be conveniently represented by a $2N \times 2N$ frequency dependent matrix, which is unitary at each frequency in the absence of MDL, such that $\mathbf{U}(z_0, z) = \mathbf{U}^{-1}(z, z_0) = \mathbf{U}^\dagger(z, z_0)$. The space and frequency evolution of $\mathbf{U}(z, z_0)$ can be expressed most generally as

$$\frac{d\mathbf{U}(z, z_0)}{dz} = i\mathbf{B}(z)\mathbf{U}(z, z_0) \tag{7}$$

$$\frac{d\mathbf{U}(z, z_0)}{d\omega} = i\mathbf{T}(z, z_0)\mathbf{U}(z, z_0), \tag{8}$$

where $\mathbf{B}(z)$ and $\mathbf{T}(z, z_0)$ are frequency dependent Hermitian matrices. Their Hermiticity is required for the unitarity of $\mathbf{U}(z, z_0)$ to be preserved at all z and ω. The multi-mode propagation problem of Eqs. (7) and (8) can be conveniently approached by extending the Pauli-matrix formalism to the multi-mode case as we show in the section that follows.

3.1. Trace orthogonal matrices and the generalized Stokes space

A generic $2N \times 2N$ Hermitian matrix $\mathbf{H}$ contains $4N^2$ entries. Of these, $2N$ are necessarily real (on the main diagonal) and the remaining $2N(2N-1)$ come in complex conjugate pairs. Thus $4N^2$ real numbers are necessary to completely define a $2N \times 2N$ Hermitian matrix. This is hence the number of basis elements in the linear space of $2N \times 2N$ Hermitian matrices. A convenient basis for this space is the one consisting of the identity matrix $\mathbf{1}$ and $4N^2 - 1$ matrices Λ_j having zero trace and satisfying the condition of trace-orthogonality

$$\frac{1}{2N}\text{Trace}\{\Lambda_i \Lambda_j\} = \delta_{i,j}, \tag{9}$$

where $\delta_{i,j}$ is the Kronecker delta function. Equation (9) can be shown to define an inner product in the space of Hermitian matrices. An example of a trace-orthogonal basis in the case of $N = 1$ is the identity with the three Pauli matrices that are used for the study of PMD [10]. The choice of a trace-orthogonal basis is never unique, and many algorithms for constructing one are possible. The Gell-Mann method [21] is one known method of construction. Its modification, which is slightly more convenient for our purposes, is described in the appendix. As should be expected, the choice of basis is immaterial for any of the physically meaningful results, and the only guideline for its selection is the simplicity of calculations. Once a basis is chosen, the matrix $\mathbf{H}$ is given by a superposition of the basis elements. A compact representation is the following

$$\mathbf{H} = \frac{1}{2N}\left(\eta_0 \mathbf{1} + \vec{\eta} \cdot \vec{\Lambda}\right), \tag{10}$$

where $\vec{\eta}$ is a vector of $4N^2 - 1$ real elements η_j, $\vec{\Lambda}$ is a vector whose elements are the matrices Λ_j, and the dot operator is implemented as

$$\vec{\eta} \cdot \vec{\Lambda} = \sum_{j=1}^{4N^2-1} \eta_j \Lambda_j. \tag{11}$$

The components of $\vec{\eta}$ can be extracted by using the trace-orthogonality of the Λ_j matrices,

$$\eta_i = \text{trace}\{\Lambda_i \mathbf{H}\} \quad \text{and} \quad \eta_0 = \text{trace}\{\mathbf{H}\}. \tag{12}$$

Note that the term proportional to η_0 represents a mode independent quantity, whereas all mode related phenomena are captured by the traceless content of $\mathbf{H}$, given by the term $\vec{\eta} \cdot \vec{\Lambda}$. The vectors $\vec{\eta}$ reside in a $4N^2 - 1$ dimensional space which is a generalization of the Stokes space used in the analysis of polarizations ($N = 1$). In what follows we will refer to vectors in this space as to *generalized Stokes vectors*, to be distinguished from the vectors $|\psi\rangle$ to which we refer as *state-vectors*. Since a projection operator $|\psi\rangle\langle\psi|$ onto a normalized state-vector $|\psi\rangle$ is a Hermitian operator, it too can be expressed in the form

$$|\psi\rangle\langle\psi| = \frac{1}{2N}\left(\mathbf{1} + \vec{\psi} \cdot \vec{\Lambda}\right). \tag{13}$$

The generalized Stokes vector $\vec{\psi}$ is obtained from $|\psi\rangle$ via Eq. (12), such that its components are given by

$$\psi_j = \text{trace}\left\{\Lambda_j |\psi\rangle\langle\psi|\right\} = \langle\psi|\Lambda_j|\psi\rangle, \tag{14}$$

identical to the usual definition of Stokes vectors in the two-polarization case. Equations (13) and (14) can be used to point out a number of interesting properties of generalized Stokes vectors. Noting that $|\langle\psi|\phi\rangle|^2 = \text{trace}\{|\psi\rangle\langle\psi|\phi\rangle\langle\phi|\}$ and substituting the right-hand-side of (13) instead of the individual projection operators, one finds that

$$|\langle\psi|\phi\rangle|^2 = \frac{1}{2N}\left(1 + \vec{\psi}\cdot\vec{\phi}\right), \tag{15}$$

where we have used the fact that trace$\{\mathbf{1}\} = 2N$, trace$\{\Lambda_j\} = 0$, as well as the identity

$$\frac{1}{2N}\text{trace}\left\{(\vec{A}\cdot\vec{\Lambda})(\vec{B}\cdot\vec{\Lambda})\right\} = \vec{A}\cdot\vec{B}, \tag{16}$$

which can readily be proven using Eqs. (9) and (11). Using $|\phi\rangle = |\psi\rangle$ in Eq. (15), one finds that $|\vec{\psi}|^2 = 2N - 1$, implying that only in the case $N = 1$, the Stokes representation of a normalized state-vector is a unit vector. If the state vectors are orthogonal to each other, Eq. (15) indicates that the product of their generalized Stokes vectors is equal to -1. Consequently, the angle α between two generalized Stokes vectors that correspond to orthogonal states satisfies $\cos(\alpha) = -(2N-1)^{-1}$, and the vectors are antipodal only in the case $N = 1$. The fact that $0 \leq |\langle\psi|\phi\rangle|^2 \leq 1$ implies that $-1 \leq \vec{\psi}\cdot\vec{\phi} \leq 2N - 1$. The left equality occurs when the state-vectors are orthogonal to each other, whereas the right equality occurs when $|\psi\rangle = |\phi\rangle$. It is evident that in the case $N > 1$ not all generalized Stokes vectors with length $\sqrt{2N-1}$ have a state vector that corresponds to them. For example, if $\vec{\phi} = -\vec{\psi}$ the product $\vec{\psi}\cdot\vec{\phi} = -2N + 1$ is smaller than -1 so that at least one of the two generalized Stokes vectors cannot represent a legitimate state. This observation can also be made directly from Eq. (13), noting that the projection operator $|\psi\rangle\langle\psi|$ has only two eigenvalues. A non-degenerate eigenvalue equal to 1 corresponds to the eigenstate $|\psi\rangle$ itself and another $2N - 1$ fold degenerate eigenvalue is equal to 0. The eigenstates of the latter eigenvalue are the $2N - 1$ state vectors that are orthogonal to $|\psi\rangle$. Thus, $\vec{\psi}$ is a legitimate representation of a state-vector only if the matrix $\vec{\psi}\cdot\vec{\Lambda}$ has a $2N - 1$ degenerate eigenvalue equal to -1 and exactly one non-degenerate eigenvalue, equal to $2N - 1$. It is obvious that only a small fraction of Stokes vectors satisfies this requirement when $N > 1$.

As discussed in Section 3.2, a relevant question concerns the relation between a generalized Stokes vector $\vec{S}$ and the eigenvalues of the matrix $\vec{S}\cdot\vec{\Lambda}$. In the case $N = 1$ the two eigenvalues are $\pm|\vec{S}|$ and the corresponding eigenvectors are those whose Stokes representation is parallel, or anti-parallel to the direction of $\vec{S}$. In the case of general N, there is no such simple relationship. The total number of eigenvalues is $2N$, and the orientation of $\vec{S}$ usually does not correspond to a legitimate state-vector. Nonetheless, the following relationship can be established. Given that $|\psi_i\rangle$ is an eigenvector of $\vec{S}\cdot\vec{\Lambda}$, thereby satisfying $\vec{S}\cdot\vec{\Lambda}|\psi_i\rangle = \theta_i|\psi_i\rangle$, where θ_i is the corresponding eigenvalue, multiplication by $\langle\psi_i|$ from the left and using (14) yields

$$\vec{S}\cdot\vec{\psi}_i = \theta_i, \tag{17}$$

implying that the largest eigenvalue corresponds to the eigenvector whose generalized Stokes representation is best aligned with the direction of $\vec{S}$, whereas the smallest eigenvalue corresponds to the eigenvector whose generalized Stokes representation is least aligned with it. A more useful relationship follows from noting that θ_i^2 is an eigenvalue of $(\vec{S}\cdot\vec{\Lambda})^2$ and hence that $\sum\theta_i^2 = \text{Trace}\left\{(\vec{S}\cdot\vec{\Lambda})^2\right\}$. Using Eq. (16), this equality produces the relation

$$2N|\vec{S}|^2 = \sum\theta_i^2. \tag{18}$$

As we shall see in what follows, this relation gives a physical meaning to the vectors that will be used to represent the matrices associated with propagation in the fiber.

3.2. Mode coupling and the mode-dispersion vector

Using Eq. (10), Eqs. (7) and (8) can be conveniently reexpressed as

$$\frac{d\mathbf{U}(z,z_0)}{dz} = \frac{i}{2N}\left(\beta_0\mathbf{1} + \vec{\beta}\cdot\vec{\Lambda}\right)\mathbf{U}(z,z_0) \tag{19}$$

$$\frac{d\mathbf{U}(z,z_0)}{d\omega} = \frac{i}{2N}\left(\tau_0\mathbf{1} + \vec{\tau}\cdot\vec{\Lambda}\right)\mathbf{U}(z,z_0), \tag{20}$$

where the dependence of β_0, τ_0, $\vec{\beta}$ and $\vec{\tau}$ on z and ω was suppressed for the simplicity of notation. Note that β_0 and τ_0 introduce a phase and group delay that is common to all the propagation modes in the fiber. These terms are therefore immaterial to our study of MD and their values can be set to zero. The vectors $\vec{\beta}$ and $\vec{\tau}$ are the two most important quantities for the study of modal dispersion in optical fibers. The first vector, $\vec{\beta}$ is a generalization of the birefringence vector, as it describes the local coupling between the various modes. If one chooses to use the basis presented in Appendix A, then the first N triplets of the components of $\vec{\beta}$ correspond to the birefringence vectors of the N individual propagation modes. Namely, $(\beta_1,\beta_2,\beta_3)$ is the birefringence vector of the first mode, $(\beta_4,\beta_5,\beta_6)$ is the birefringence vector of the second mode and so on. The remaining $4N^2 - 3N - 1$ terms of $\vec{\beta}$ represent coupling between spatial modes, or the difference in the spatial modes' propagation vectors. The vector $\vec{\tau}$ is the direct generalization of the PMD vector [10] and it captures the essentials of MD.

Equations (19) and (20) can be used to establish the concatenation rule satisfied by the MD vector $\vec{\tau}$. To this end let us consider an assembly of two concatenated fiber sections characterized by the two matrices $\mathbf{U}_1$ and $\mathbf{U}_2$. The matrix describing their concatenation is $\mathbf{U} = \mathbf{U}_2\mathbf{U}_1$. One may express the combined MD vector $\vec{\tau}$ of the two fibers as

$$\vec{\tau}\cdot\vec{\Lambda} = -2Ni\frac{d\mathbf{U}}{d\omega}\mathbf{U}^\dagger = -2Ni\frac{d\mathbf{U}_2}{d\omega}\mathbf{U}_2^\dagger + \mathbf{U}_2\left(-2Ni\frac{d\mathbf{U}_1}{d\omega}\mathbf{U}_1^\dagger\right)\mathbf{U}_2^\dagger = \vec{\tau}_2\cdot\vec{\Lambda} + \mathbf{U}_2\left(\vec{\tau}_1\cdot\vec{\Lambda}\right)\mathbf{U}_2^\dagger, \tag{21}$$

where $\vec{\tau}_1$ and $\vec{\tau}_2$ are the MD vectors of the first and second fiber section, respectively. The concatenation rule (21) can be stated as follows: The joint MD vector equals the MD vector of the first section rotated by the coupling matrix of the second section and supplemented by the MD vector of the second section alone. This description is identical to that used in the case of PMD [10], although the word rotation is applied more loosely in the present case, indicating an operation that does not change the length of a generalized stokes vector. To see that this is indeed the case with the second term in (21), one may define $\vec{\tau}_1^{(R)}\cdot\vec{\Lambda} = \mathbf{U}_2\left(\vec{\tau}_1\cdot\vec{\Lambda}\right)\mathbf{U}_2^\dagger$ and take advantage of identity (16) with $\vec{A} = \vec{B}$ to show that

$$\left|\vec{\tau}_1^{(R)}\right|^2 = \frac{1}{2N}\text{trace}\left\{\left[\mathbf{U}_2\left(\vec{\tau}_1\cdot\vec{\Lambda}\right)\mathbf{U}_2^\dagger\right]^2\right\} = |\vec{\tau}_1|^2, \tag{22}$$

where we made use of the fact that the trace of a product is invariant under cyclic permutations. It is instructive to consider the case in which $\mathbf{U}_1$ represents propagation from z_0 to z whereas $\mathbf{U}_2$ represents an incremental evolution between the points z and $z + dz$. Using Eq. (19) one finds that to first order in dz, $\mathbf{U}_2 = \mathbf{1} + (i/2N)\vec{\beta}(z,\omega)\cdot\vec{\Lambda}dz$. Substituting this into (21) produces

$$\frac{\partial}{\partial z}\left(\vec{\tau}\cdot\vec{\Lambda}\right) = \vec{\beta}_\omega\cdot\vec{\Lambda} + \frac{i}{2N}\left[\left(\vec{\beta}\cdot\vec{\Lambda}\right)\left(\vec{\tau}\cdot\vec{\Lambda}\right) - \left(\vec{\tau}\cdot\vec{\Lambda}\right)\left(\vec{\beta}\cdot\vec{\Lambda}\right)\right], \tag{23}$$

with $\vec{\beta}_\omega = \partial\vec{\beta}/\partial\omega$. This expression reduces to Eq. (6.15) of [10] in the case of $N = 1$. The solution to Eq. (23) is formally identical to Eq. (7.22) of [10],

$$\vec{\tau}(L) = \int_0^L \mathbf{R}(L,z)\vec{\beta}_\omega dz \tag{24}$$

where $\mathbf{R}(L,z)$ is a $D \times D$ norm-preserving matrix accounting for the evolution of the infinitesimal contribution $\vec{\beta}_\omega(z)\mathrm{d}z$ from z to the fiber end at $z = L$ and related to $\mathbf{U}(L,z)$ by

$$\mathbf{U}(L,z)\left(\vec{\beta}_\omega \cdot \vec{\Lambda}\right)\mathbf{U}^\dagger(L,z) = \left(\mathbf{R}(L,z)\vec{\beta}_\omega\right) \cdot \vec{\Lambda}. \tag{25}$$

The physical meaning of $\vec{\tau}$ is provided by Eqs. (17), (18). Indeed, if we use $\vec{S} = \vec{\tau}/(2N)$, then the eigenvalues θ_i coincide with the group delays of the individual principal states t_i and Eq. (17) becomes $t_i = \vec{\tau} \cdot \vec{\psi}_i/(2N)$, thus showing that these are bounded between $-\tau\sqrt{2N-1}/(2N)$ and $\tau\sqrt{2N-1}/(2N)$. On the other hand, Eq. (18) becomes

$$\tau^2 = 2N \sum_{i=1}^{2N} t_i^2, \tag{26}$$

which gives the magnitude of the MD vector as proportional to the sum of the square group delays. For $N = 1$, τ is equal to the differential group delay $|t_2 - t_1|$, as expected.

An alternative and much simpler formulation notationally follows by defining a cross-product operation in the generalized Stokes space. The Hermiticity and the zero-trace property of the matrices Λ_i imply that $i(\Lambda_i\Lambda_j - \Lambda_j\Lambda_i)/(2N)$ is also a traceless Hermitian matrix. Hence it can be expanded as $i(\Lambda_i\Lambda_j - \Lambda_j\Lambda_i)/(2N) = \sum_k f_{i,j,k}\Lambda_k$, with $f_{i,j,k}$ being a set of real-valued structure constants,

$$f_{i,j,k} = \frac{i}{(2N)^2}\mathrm{trace}\left\{\Lambda_k\left(\Lambda_i\Lambda_j - \Lambda_j\Lambda_i\right)\right\}. \tag{27}$$

Using this representation we define a generalized cross-product as

$$\vec{A} \times \vec{B} = \sum_{i,j,k} f_{i,j,k}A_iB_j\,\vec{e}_k \tag{28}$$

where $\vec{e}_i$ are the orthogonal unit-length basis vector of the D-dimensional generalized Stokes space [24]. Equation (28) can be used to re-express the second term on the right-hand-side of Eq. (23) as $(\vec{\beta} \times \vec{\tau}) \cdot \vec{\Lambda}$, so that Eq. (23) simplifies to

$$\frac{\partial \vec{\tau}}{\partial z} = \vec{\beta}_\omega + \vec{\beta} \times \vec{\tau}, \tag{29}$$

generalizing the familiar equation for the evolution of the PMD vector [10] to the multimode case. By differentiating $\vec{S} = \langle\psi(0)|\mathbf{U}(z,0)^\dagger\vec{\Lambda}\mathbf{U}(z,0)|\psi(0)\rangle$ with respect to z and ω, and by using Eqs. (19) and (20), the evolution of the generalized Stokes vector in space and frequency can be expressed as

$$\frac{\partial \vec{S}}{\partial z} = \vec{\beta} \times \vec{S}, \qquad \frac{\partial \vec{S}}{\partial \omega} = \vec{\tau} \times \vec{S}, \tag{30}$$

generalizing, once again, the corresponding equations obtained in the single-mode fiber case.

4. Statistical analysis

In order to account for the statistical properties of MD, we assume that the distribution of the mode coupling vector $\vec{\beta}$ can be adequately modeled as white Gaussian noise. Although this assumption is not strictly satisfied even in single-mode optical fibers [22], it was shown to produce a very accurate description of PMD when the overall fiber length is significantly larger than the correlation length of the local birefringence. In principle, the white noise model

is viable whenever the correlation length of the mode coupling vector is very small relative to the overall system length. Since the source of perturbations (geometrical distortions and mechanical stress) are similar to those experienced by single-mode fibers, it is expected that the correlation length of $\vec{\beta}(z)$ is of a similar order of magnitude as the correlation length of the polarization birefringence vector in single-mode fibers. Using this assumption, it is quite obvious that the MD vector $\vec{\tau}$ is Gaussian and evolves as a Wiener process in z, which is the continuous limit of a random walk. This conclusion can be readily drawn from Eq. (24) which shows that $\vec{\tau}$ is the sum of rotated and statistically independent increments of $\vec{\beta}_\omega$. Exploiting similar arguments to those made in the analysis of PMD statistics [23], one can conclude that the vector $\vec{\tau}(\omega, z)$ experiences a D dimensional isotropic random walk, hence it is Gaussian distributed and its modulus τ obeys a chi distribution. An important consequence of Eq. (24) is that the mean-square MD value $\langle \tau^2 \rangle$ is equal to $\langle \tau^2 \rangle = \gamma^2 z$, where γ^2 is defined through the relation $\langle \vec{\beta}_\omega(z) \cdot \vec{\beta}_\omega(z') \rangle = \gamma^2 \delta(z - z')$. In the case of partial coupling, $\vec{\tau}$ is still Gaussian if $\vec{\beta}(z)$ is a white noise process, yet its distribution may not be isotropic, and the equation $\langle \tau^2 \rangle = \gamma^2 z$ still rigorously applies. In this case, numerical modeling shows that τ approximately obeys a chi distribution with an effective number of degrees of freedom $D_{\mathrm{eff}} < D$.

The derivation of the autocorrelation functions Eqs. (4) and (5) is performed most simply by using the relations introduced in Eq. (29). Just as in the procedure followed in the study of PMD [19], we switch to a rotating reference frame, where, after a first-order expansion in frequency, $\vec{\beta}$ is replaced by $\omega \vec{\beta}_\omega$ with ω denoting the offset from the central frequency of the propagating signal. The spatial dependence of $\vec{\beta}_\omega$ is modeled as white Gaussian noise. The customary procedure in this case is to represent the integral of $\vec{\beta}_\omega$ over the propagation distance as an isotropic D-dimensional Wiener process $\vec{W}(z)$. The increments of $\vec{W}(z)$ are characterized by $\langle \mathrm{d}W_i(z)\mathrm{d}W_j(z) \rangle = D^{-1}\gamma^2 \delta_{i,j}\mathrm{d}z$, where γ is a constant proportional to the mean-square magnitude of the generalized birefringence, such that $\langle \tau^2(z) \rangle = \gamma^2 z$. Equation (29) can be re-expressed as a stochastic differential equation

$$\mathrm{d}\vec{\tau} = \mathrm{d}\vec{W} + \omega \mathrm{d}\vec{W} \times \vec{\tau}, \tag{31}$$

which is to be interpreted in the Stratonovich sense. The corresponding Ito equation involves an additive correction term $-(\gamma^2 \omega^2 / D)\vec{\tau}\mathrm{d}z$, which is obtained after some algebra. From this point on, considering that our cross-product possesses the property $\vec{\tau} \cdot (\mathrm{d}\vec{W} \times \vec{\tau}) = 0$, the derivation of the autocorrelation functions is identical to the one presented in [19]. The only modification is that the term D generalizes the factor 3 appearing in the corresponding term in [19].

Finally, we note that the scaling with the number of degrees of freedom is such that the shape of the autocorrelation functions is uniquely determined by $\langle \tau^2 \rangle / D$, a quantity that has the meaning of the mean square MD parameter per dimension in the extended Stokes space. Polarization mode dispersion in single-mode fibers is a special case corresponding to $D = 3$.

4.1. Numerical simulations

In the numerical simulations a fiber is modeled as an assembly of N_s wave-plates, such that the k-th element is characterized by a generalized real-valued D-dimensional birefringence vector $\vec{b}_{\mathrm{det}} + \vec{b}_k$, where $\vec{b}_{\mathrm{det}}$ describes the deterministic birefringence and $\vec{b}_k$ accounts for perturbation-induced birefringence. In the simulation of quasi-degenerate modes $\vec{b}_{\mathrm{det}} = 0$, whereas in the study of non-degenerate modes reported in Fig. 4 $\vec{b}_{\mathrm{det}}$ is used to tune the degree of deterministic coupling between different groups of quasi-degenerate modes. Since $\vec{b}_{\mathrm{det}}$ describes wave-vector mismatch between different modes, as discussed in Appendix A, only the last $N - 1$ components of $\vec{b}_{\mathrm{det}}$ can be nonzero. Expanding $\vec{b}_k$ to first order in the optical frequency, and omitting the frequency-independent zeroth-order term (whose effect on the statistics of modal

dispersion is immaterial) we write $\vec{b}_k = \omega \vec{b}'_k$, where ω denotes the offset from the central frequency of the transmitted signal. The transfer matrix of the k-th section is therefore given by $\mathbf{U}_k = \exp\left[i(\vec{b}_{\text{det}} + \omega \vec{b}'_k) \cdot \vec{\Lambda}/(2N)\right]$. The frequency independent part of $\vec{b}_k$ was verified to have no effect on the statistics of MD and hence it was omitted in the simulations. A similar omission of the frequency independent part of the random birefringence is customary in the study of polarization effects and it is known to have no consequences on the statistics when characterizing linear propagation problems. The vectors $\vec{b}'_k$ are selected independently of one another and with each having D statistically independent zero-mean Gaussian components.

We start from discussing the case of quasi-degenerate modes, when $\vec{b}_{\text{det}} = 0$. In this case, using Eq. (21), the average square-modulus of the MD vector can be shown to be $\langle \tau^2 \rangle = N_s \langle |\vec{b}'_k|^2 \rangle$. To link this notation to the one used in prior sections we note that $\vec{b}_k$ is equivalent to $\vec{\beta}(z_k)\Delta z_k$ with z_k being the position of the k-th fiber increment and with Δz_k being its length.

In the case of full coupling, all components of the vector $\vec{b}'_k$ are identically distributed and their variance is taken to be $\sigma^2 = \langle \tau^2 \rangle/(N_s D)$. While the modeling of the full coupling case is relatively straightforward, the case of partial coupling is significantly more involved. Most importantly, the regime of partial coupling cannot be captured unequivocally in terms of a single parameter. Recent experimental work [20] characterized mode coupling between modes i and j in terms of the fraction of power that is coupled into mode j at the end of the link when mode i is excited. While this is a practical and useful parameter for quantifying mode coupling, it may not provide a unique characterization of the MD statistics, as the same coupling can be generated in a variety of ways. While a detailed description of the partial coupling regime in multi-mode fibers requires further experimental investigation, we model this situation in two ways. If partial coupling is caused by the phase mismatch between non-degenerate modes, then the magnitude of $\vec{b}_{\text{det}}$ controls the degree of coupling. The results reported in Fig. 4 refer to this situation. Otherwise, in the case of quasi-degenerate modes, partial coupling can only be caused by an asymmetry of the coupling strength itself. This situation is modeled by attaching different values to the variance of different components of $\vec{b}'_k$. In particular, using our choice of the matrices constructing $\vec{\Lambda}$ (see Appendix A), the first $3N$ components of $\vec{b}'_k$ are responsible for polarization coupling within the same spatial mode, whereas the remaining components are responsible for coupling different spatial modes with one another. Thus the degree of inter-modal coupling is controlled by varying the variance σ_c^2 of the last $D - 3N$ components of $\vec{b}'_k$, as compared to the variance of the first $3N$ components, which we denote by σ_p^2. In this case the average square modulus of the MD vector is given by $\langle \tau^2 \rangle = [3N\sigma_p^2 + (D - 3N)\sigma_c^2]N_s$.

5. Conclusions

Inspired by the analogy between MD and PMD, we developed a formalism to characterize modal dispersion in multimode fibers. The proposed formalism is based on the definition of a set of $D = 4N^2 - 1$ matrices — which are the generalization of the Pauli spin matrices — as a basis for the expansion of the generalized Jones matrix $\mathbf{U}$ describing the propagation of N spatial modes. In the absence of mode-dependent loss, MD is characterized by a real D-dimensional MD vector $\vec{\tau}$, whose components are the coefficients of the expansion of the traceless part of the matrix $\mathbf{T} = -i(d\mathbf{U}/d\omega)\mathbf{U}^\dagger$. The eigenstates of such a matrix are generalized principal states and its eigenvalues are the corresponding group delays; the square length of $\vec{\tau}$ is the sum of the square group delays. In the regime of strong mode coupling the vector $\vec{\tau}$ is Gaussian distributed and its modulus follows a chi distribution with D degrees of freedom. We found that to an excellent degree of approximation, the same distribution, yet with a smaller number of degrees of freedom, also characterizes the largest differential group delay, that is the difference between the delays of the fastest and the slowest principal states. Finally, we studied the frequency

autocorrelation of the MD vector, of the square modulus of the MD vector and of the largest differential group delay. Extensive numerical simulations were used to check the analytical results, showing good agreement in all cases.

A. Algorithm to build the Λ matrices

For $N = 1$ the matrices Λ_i coincide with the Pauli matrices, namely

$$\Lambda_1 = \sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ \Lambda_2 = \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \Lambda_3 = \sigma_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{32}$$

For $N > 1$ the algorithm which gives the $4N^2 - 1$ matrices proceeds as follows.

From Λ_1 to Λ_{3N}: Each 2×2 block in the main diagonal, from the leftmost to the rightmost, is sequentially set to either one of the three Pauli matrices, while the other elements are set to zero. This gives $3N$ matrices. The trace of such matrices squared is 2 and hence a normalization coefficient $\sqrt{N}$ is required.

From Λ_{3N+1} to Λ_{4N^2-N}: Each element outside the 2×2 blocks in the main diagonal is sequentially set to either 1 or i and the symmetric element to 1 or $-i$ respectively, while the other elements are set to zero. This gives $4N^2 - 4N$ matrices. The trace of such matrices squared is 2 and hence a normalization coefficient $\sqrt{N}$ is required.

From Λ_{4N^2-N+1} to Λ_{4N^2-1}: We denote these matrices as Λ_{4N^2-N+n}, with $n = 1, \ldots, N-1$. The 2×2 blocks in the main diagonal from the first to the n-th are set to σ_0 and the $(n+1)$-th to $-n\sigma_0$, while the other elements are set to zero. The trace of such matrix squared is $2(n^2 + n)$ and hence a normalization coefficient $\sqrt{N/(n^2+n)}$ is required.

It is apparent from the definition that the matrices $\Lambda_1 \ldots \Lambda_{3N}$ do not couple different spatial modes. Correspondingly, the components $\beta_1 \ldots \beta_{3N}$ of the generalized birefringence vector $\vec{\beta}$ account for polarization-mode coupling within the same spatial mode. The matrices Λ_{3N+1} to Λ_{4N^2-N} couple individual polarization modes of different spatial modes, whose magnitude is defined by the corresponding components of $\vec{\beta}$. Finally, the matrices Λ_{4N^2-N} to Λ_{4N^2-1} describe wave-vector mismatch between different spatial modes.

As an example, we list below the Λ_i matrices for the case $N = 2$.

$$\Lambda_1 = \sqrt{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ \Lambda_2 = \sqrt{2} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ \Lambda_3 = \sqrt{2} \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\Lambda_4 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \ \Lambda_5 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \ \Lambda_6 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix},$$

$$\Lambda_7 = \sqrt{2} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ \Lambda_8 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \ \Lambda_9 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\Lambda_{10} = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \ \Lambda_{11} = \sqrt{2} \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ \Lambda_{12} = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix},$$

$$\Lambda_{13} = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ \Lambda_{14} = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix}, \ \Lambda_{15} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Acknowledgment