Optimal Launch States for the Measurement of Principal Modes in Optical Fibers

Ioannis Roudas , *Member, IEEE*, Jaroslaw Kwapisz, and Daniel A. Nolan, *Fellow, OSA*

Abstract—**Modal dispersion characterization of multimode optical fibers can be performed using the recently proposed mode-dependent signal delay method. This method consists of sending optical pulses using different combinations of modes through the multimode optical fiber and measuring the mode group delay at the fiber output. From these measurements, it is possible to estimate the modal dispersion vector, the principal modes, and their corresponding differential mode group delays. In this paper, we revise and extend the theoretical framework of the mode-dependent signal delay method to include the impact of receiver noise and mode-dependent loss. We compute optimal launch modes, minimizing the noise error in the estimation of the fiber modal dispersion vector. We show that, for a 40-mode fiber, the electronic signal-to-noise ratio is improved asymptotically by almost 6 dB compared to conventional mode combinations.**

Index Terms—**Modal dispersion, multimode fiber characterization.**

I. Introduction

INTERNET traffic is expected to grow steadily in the near future [1]. For instance, Cisco predicts a 24% compound annual growth of global data traffic through the Internet from 2016 to 2021 [2]. If this trend persists over longer periods of time, it could eventually lead to a capacity shortage in the global fiber-optic network [3].

To address this challenge, researchers have been considering for some time the introduction of new fiber technologies that can support petascale data traffic per link in a cost-efficient way. For instance, it is possible to increase link capacity by using spatial division multiplexing (SDM), i.e., parallel transmission of optical data streams over disjointed spatial paths [4] provided by multimode and multicore optical fibers (jointly abbreviated below by the composite acronym SDM MMFs) [5]. So far, it has been shown, both theoretically and experimentally, that strongly-coupled, single-mode, homogeneous multicore fibers exhibit a slight performance advantage over single-mode fiber

(SMF) bundles [6]–[8]. Whether this argument alone is sufficient for the adoption of such optical fibers by the telecommunications market remains to be seen.

It is assumed here that SDM MMFs will be eventually adopted in long-haul optical communications systems. A distinct feature of SDM MMFs for long-haul applications is that their modal dispersion (MD) should be very low, ideally comparable to the levels of polarization-mode dispersion (PMD) of SMFs, in order to facilitate digital signal processing at the coherent optical receiver [4].

By analogy with PMD [9]–[11], MD in long SDM MMFs can be described by a set of propagation modes called principal modes (PMs) and by their corresponding differential mode group delays (DMGDs) compared to the average mode group delay [12]. These quantities can be geometrically represented by a vector in a generalized Stokes space called MD vector [13], [14], which is a direct extension of the PMD vector in the conventional Stokes space [9]–[11].

Taking further advantage of the similarity between MD and PMD, it is possible to modify previously-proposed PMD measurement techniques [11] and use them for MD characterization [15]–[17]. For instance, the polarization-dependent signal delay method can be used for the measurement of the PMD vector of SMFs [18]. The recently-proposed mode-dependent signal delay method [19] is a generalization of the polarization-dependent signal delay method that can be used for the measurement of the MD vector of SDM MMFs. Namely, it relies on the determination of the components of the MD vector by launching optical pulses corresponding to different combinations of modes at the fiber input and measuring the corresponding group delays at the fiber output [19]. From the MD vector, one can construct a Hermitian matrix, called the group-delay operator, and determine the PMs and the DMGDs from its eigenvectors and eigenvalues, respectively [13], [14].

An important question that is left unanswered in previous articles on the mode-dependent signal delay method [19], [20] is which launch modes must be used to measure the MD vector. Let N be the number of spatial and polarization modes in the SDM MMF under test. Then, the dimensionality of the generalized Stokes space is $N^2 - 1$ [13], [14]. We need to choose $N^2 - 1$ different combinations of launch modes to determine the $N^2 - 1$ components of the MD vector. A set of $N^2 - 1$ launch states corresponding to $N^2 - 1$ orthonormal vectors in Stokes space would constitute the best coordinate system for conducting such measurements. For $N > 2$, however, it is impossible to find $N^2 - 1$ launch mode combinations corresponding to $N^2 - 1$

orthonormal vectors in the generalized Stokes space due to the incomplete coverage of the Poincaré sphere with valid states [13], [14].

Milione *et al.* do not address this issue [19], while Yang and Nolan [20] propose a set of modes that is a generalization into higher dimensions of the linear horizontal, linear 45 deg, and right-circular states of polarization (SOPs) used for measurements conducted in the conventional three-dimensional Stokes space. A weakness of the mode set proposed by [20] is that, for $N > 2$, it leads to suboptimal performance in the presence of receiver noise, as shown in Section V.

In the present paper, we analyze the impact of noise on the MD vector characterization process performed in the mode-dependent signal delay method. Our analysis reveals that using a set of launch modes corresponding to an oblique vector set in the generalized Stokes space always magnifies the error in the estimation of the MD vector. Therefore, we should seek launch states that correspond to maximally-orthogonal Stokes vectors in order to minimize this error. In the sections that follow, we propose two numerical optimization algorithms based on the gradient descent method [21] that search the generalized Stokes space for $N^2 - 1$ maximally-orthogonal vectors corresponding to feasible launch states.

Furthermore, the presence of mode-dependent loss (MDL) is not considered in the original articles on the mode-dependent signal delay method [19], [20]. This is a significant omission in the sense that long SDM MMFs always exhibit a certain amount of MDL. In the present paper, we show that the mode-dependent signal delay method can be modified to characterize both MD and MDL simultaneously.

In the following, we derive, from first principles, the fundamental equations for the determination of the MD and MDL vectors by using the mode-dependent signal delay method in the presence of additive white Gaussian noise, MD, and MDL (Sections II-A–II-I). Practical considerations that might affect the accuracy of the mode-dependent signal delay method are discussed in Section II-J. The remainder of this paper is devoted to the description of two optimization algorithms based on the gradient descent method for the selection of maximally-orthogonal launch states (Sections III and IV, respectively). Using these algorithms, we compute optimal sets of launch modes for up to $N = 40$ that maximize the signal-to-noise ratio (SNR) at the direct-detection receiver and enhance the accuracy of the mode-dependent signal delay method (Section V). For example, for a 40-mode SDM MMF, we show that the optimal mode combinations improve the noise performance of the mode-dependent signal delay method by almost 6 dB compared to the set of modes proposed by Yang and Nolan [20]. We also compare the noise performance of the proposed optimal mode combinations to vector sets often used for measurements in quantum mechanics, i.e., symmetric, informationally complete, positive operator valued measure (SIC-POVM) vectors [22] and vectors selected from mutually unbiased bases (MUBs) [23] (see Appendix A). We show that the proposed optimal mode combinations exhibit superior noise performance asymptotically by 3 dB compared to the aforementioned vector sets.

II. MATHEMATICAL MODEL

A. Fundamental Concepts

An intuitive way to quantify the modal dispersion of an optical fiber is through the time-of-flight of an optical pulse traveling along the fiber. The optical pulse experiences different delays depending on the combination of modes that are excited at the fiber input, the group velocity differences among the propagation modes, and the mode coupling at various points inside the fiber. The received pulse is a mixture of a multitude of components that arrive at the receiver at slightly different times and interfere constructively or destructively. The extraction of information regarding the fiber modal dispersion from pulse delays, exclusively, constitutes the cornerstone of the mode-dependent signal delay method.

In the absence of MDL (see Section II-F), an arbitrary launch mode can always be written as a linear combination of principal modes at the fiber input. Furthermore, the group delay of a narrowband optical pulse at the fiber output can be written as a weighted superposition of the group delays experienced by the principal modes. The weights of the superposition are functions of the excitations of the input principal modes. Formally, this weighted superposition can be written in a concise form in Stokes space as the dot product between the input MD vector representing the MD of the optical fiber and a unit Stokes vector representing the launch combination of modes (see expression (12)).

The basic idea of the mode-dependent signal delay method is to identify the components of the MD vector. This can be accomplished by launching optical pulses at the fiber input corresponding to different combinations of modes and measuring the corresponding group delays at the fiber output [19]. Assume that different pulses excite mode combinations corresponding to vectors that linearly span Stokes space. Then, the group delays experienced by different pulses are enough to recover the input MD vector.

If the above vectors are linearly independent, this corresponds to a decomposition of the MD vector into the basis of the Stokes vectors representing the launch combinations of modes. We can write the components of the Stokes vectors in the form of a matrix, the coefficient matrix of the linear system (see expression (18)). We can also write the group delays corresponding to different mode combinations as a column vector (see expression (19)). The components of the MD vector can then be recovered by solving a set of linear equations (see expression (21)).

If the Stokes vectors representing the mode combinations are pairwise orthogonal, the solution of the aforementioned set of linear equations is less affected by the unavoidable presence of random perturbations in the measurements (e.g., thermal noise, errors in the settings of the mode converter). Otherwise, matrix inversion leads to error amplification.

We follow a similar procedure to the one described above in order to measure the MDL of an optical fiber (see Section II-I). For MDL characterization, we perform measurements of the average output power of CW optical waves. A continuous

optical wave can always be decomposed into a superposition of principal attenuation modes at the fiber input. The attenuation of the optical wave can be written as a weighted superposition of the attenuations experienced by the principal attenuation modes. The weights of the superposition are functions of the excitations of the input principal attenuation modes. Formally, this weighted superposition can be written in a concise form in Stokes space as the dot product between the input MDL vector, representing the MDL of the optical fiber, and a unit Stokes vector representing the launch combination of modes (see expression (56)).

In summary, the mode-dependent signal delay method can be used to measure both MD and MDL. In the joint presence of MD and MDL, we use a more sophisticated, multi-step version of the mode-dependent signal delay method, involving two separate sets of measurements, for MDL and MD, respectively, and an intermediate MDL equalization phase (see Section II-I). First, we carry out MDL characterization by sending CW light into the fiber under test and measuring the average output power, as described above. Measurements for different launch mode combinations enable us to retrieve the individual components of the MDL vector. Once the MDL vector is known, we can optically compensate for the optical fiber's MDL at the fiber input by adjusting the settings of the mode converter. Finally, the MD vector of the MDL-equalized optical fiber can be determined by launching optical pulses at the fiber input corresponding to different combinations of modes and measuring the corresponding group delays at the fiber output, as described in the beginning of this subsection. In both sets of measurements, for MDL and MD characterization, we can use the same set of launch modes corresponding to maximally-orthogonal Stokes vectors in order to minimize noise errors.

On a final note, it is worth explaining the motivation behind using the generalized Stokes formalism for the analysis of the experimental measurements provided by the mode-dependent signal delay method. The familiar three-dimensional (3D) Stokes space is traditionally used to provide a geometrical representation of SOPs in terms of real vectors. This is useful for visualization of the transformations of SOPs by optical systems. For instance, the spatial evolution of the SOP of a monochromatic optical plane wave traveling through a birefringent SMF can be graphically depicted by a trajectory on the surface of the Poincaré sphere. The Stokes space formalism may be extended to higher dimensions [13], [14] to describe geometrically mode combinations and their evolution during propagation through "modally-birefringent" SDM MMFs, but then the intuitive visual appeal of the conventional 3D Stokes representation is lost. Nevertheless, from a modeling perspective, there is still an incentive for using the generalized Stokes space instead of the generalized Jones space. The reason is that measurable quantities provided by direct-detection receivers can be conveniently expressed as dot products of generalized Stokes vectors. Therefore, the use of the generalized Stokes space enables us to perform analytical calculations solely in terms of real vectors, avoiding to employ complex matrices in the generalized Jones space.
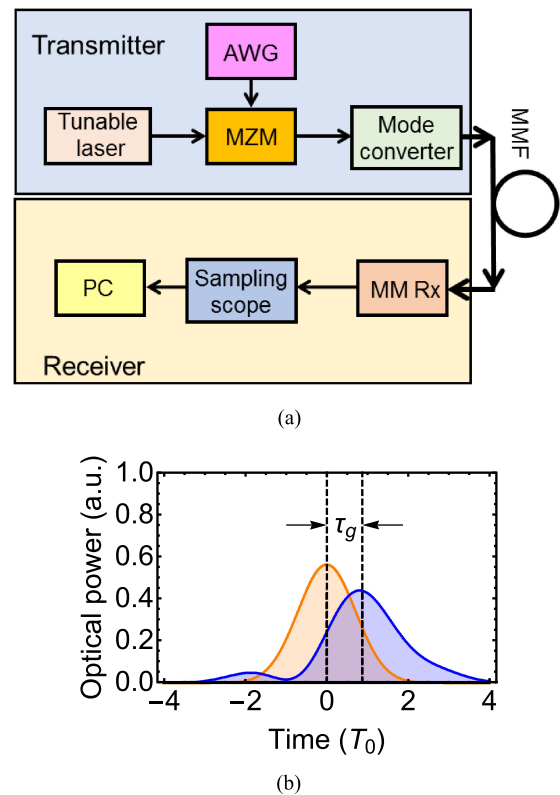

(a)

(b)

Fig. 1. (a) Experimental setup for MD characterization of SDM MMFs using the mode-dependent signal delay method (Abbreviations: AWG: arbitrary waveform generator, MZM: Mach-Zehnder modulator, MMF: multimode fiber, MM Rx: multimode receiver, PC: computer); (b) Input Gaussian pulse with unit energy (orange line) and output pulse (blue line). The origin of the time axis coincides with the center of the input pulse. (Symbols: T_0 = half-width at 1/e power point [24]; τ_g = group delay of the output pulse).

B. Experimental Process

The experimental setup used in the mode-dependent signal delay method is shown in Fig. 1(a). The transmitter consists of a tunable laser externally modulated by a Mach-Zehnder modulator (MZM) using an arbitrary waveform generator (AWG). This configuration provides narrowband, Gaussian, transform-limited, optical pulses with adjustable duration and repetition rates. The laser scans the whole frequency band of interest in steps larger than the pulse bandwidth. The purpose of the mode converter is to generate arbitrary spatial and polarization mode combinations. Applying appropriate mode excitations, one can launch optical pulses which experience different propagation delays. After the SDM MMF under test, the optical pulses are detected, sampled using a real-time oscilloscope, and stored in a PC. Off-line processing can be used for the evaluation of the PMs and the DMGDs at each frequency.

Indicative drawings of the input and output pulses are shown in Fig. 1(b) (orange and blue lines, respectively). In the following subsections, we link the group delay τ_g to the MD vector and the unit Stokes vector representing the launch combination of modes.

C. Basic Notation

Throughout this article, we use the notation conventions introduced by [10]: Dirac's bra-kets refer to unit Jones vectors. Unit Stokes vectors are denoted by carets, while non-unit Stokes vectors are denoted by arrows. Matrices are designated by uppercase boldface letters.

We can write any $N \times N$ matrix $\mathbf{M}$ as a linear superposition of the identity matrix $\mathbf{I}$ and the $N^2 - 1$ generalized Gell-Mann matrices $\mathbf{\Lambda}_i, i = 1, \ldots, N^2 - 1$ [14]. We denote by $\mathbf{\Lambda}$ the Gell-Mann column vector

$$\mathbf{\Lambda} := [\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_{N^2-1}]^T, \tag{1}$$

where the superscript T denotes the transpose of a matrix.

The unit Stokes vector $\hat{s}$ corresponding to the unit Jones vector $|s\rangle$ is defined as [14]

$$\hat{s} := C_N \langle s | \mathbf{\Lambda} | s \rangle, \tag{2}$$

where C_N denotes the normalization coefficient [14]

$$C_N := \sqrt{N / [2(N-1)]}. \tag{3}$$

For each Jones vector $|s\rangle$, we can define the associated projection operator $|s\rangle\langle s|$, which represents a mode filter, i.e., the equivalent of a polarizer in the two-dimensional case. The projection operator can be expressed in terms of the identity matrix and the generalized Gell-Mann matrices [14]

$$|s\rangle\langle s| = \frac{1}{N}\mathbf{I} + \frac{1}{2C_N}\hat{s} \cdot \mathbf{\Lambda}. \tag{4}$$

In (4), we used the dot product of the Stokes vector $\hat{s}$ with the Gell-Mann vector $\mathbf{\Lambda}$. This is defined as the sum of the products of the corresponding entries

$$\hat{s} \cdot \mathbf{\Lambda} := \sum_{i=1}^{N^2-1} s_i \mathbf{\Lambda}_i \tag{5}$$

Finally, we will use, without proof, a relationship between the inner product in Stokes space and the inner product in Jones space derived in [14]:

$$\hat{s}_j \cdot \hat{s}_k = 2C_N^2 \left[|\langle s_j | s_k \rangle|^2 - \frac{1}{N} \right]. \tag{6}$$

D. MD Vector Definition

First, we consider the ideal case of an N-mode SDM MMF with negligible MDL. The fiber transfer function can be described by a generalized Jones unitary matrix $\mathbf{U}(\omega)$. We define the input group-delay operator $i\mathbf{U}^\dagger(\omega)\mathbf{U}_\omega(\omega)$, where the subscript ω denotes differentiation with respect to the angular frequency and a raised dagger denotes the adjoint matrix.

The input group-delay operator can be represented in the basis of the identity matrix and the generalized Gell-Mann matrices

$$i\mathbf{U}^\dagger(\omega)\mathbf{U}_\omega(\omega) := \tau_0(\omega)\mathbf{I} + \frac{1}{2C_N}\vec{\tau}_s(\omega) \cdot \mathbf{\Lambda}, \tag{7}$$

where $\tau_0(\omega)$ is the average group delay and $\vec{\tau}_s(\omega)$ is the input MD vector [14].

The eigenvalues and eigenvectors of the operator $1/(2C_N)\vec{\tau}_s(\omega) \cdot \mathbf{\Lambda}$ are the DMGDs $\tau_i(\omega)$ and the input PMs $|p_i(\omega)\rangle, i = 1, \ldots, N$. We can write the eigenvalue equation

$$\frac{1}{2C_N}\vec{\tau}_s(\omega) \cdot \mathbf{\Lambda} |p_i(\omega)\rangle = \tau_i(\omega)|p_i(\omega)\rangle. \tag{8}$$

E. Information Provided by the Input MD Vector

Before we delve further into the mode-dependent signal delay method, it is worth investigating whether the MD vector $\vec{\tau}_s(\omega)$ and the fiber transfer matrix $\mathbf{U}(\omega)$ are really equivalent representations of the optical fiber modal dispersion.

From the definition of the input MD vector (7), setting the average group delay $\tau_0 = 0$ for simplicity, we obtain

$$i\mathbf{U}^\dagger(\omega)\mathbf{U}_\omega(\omega) = \frac{1}{2C_N}\vec{\tau}_s(\omega) \cdot \mathbf{\Lambda}. \tag{9}$$

Acting on both sides of this expression from the left with $-i\mathbf{U}(\omega)$ yields a $N \times N$ homogeneous system of coupled first-order ordinary differential equations

$$\mathbf{U}_\omega(\omega) = \mathbf{U}(\omega)\left[-\frac{i}{2C_N}\vec{\tau}_s(\omega) \cdot \mathbf{\Lambda}\right]. \tag{10}$$

Assume that $\vec{\tau}_s(\omega)$ is constant in the interval $[\omega_0, \omega_0 + \delta\omega]$. Then, the solution of the above system in matrix form is

$$\mathbf{U}(\omega_0 + \delta\omega) = \mathbf{U}(\omega_0)\exp\left[-\frac{i}{2C_N}\vec{\tau}_s(\omega_0) \cdot \mathbf{\Lambda}\delta\omega\right]. \tag{11}$$

Notice that knowledge of $\vec{\tau}_s(\omega_0)$ is insufficient to fully determine the fiber transfer matrix $\mathbf{U}(\omega_0 + \delta\omega)$. This would require knowledge of $\mathbf{U}(\omega_0)$ as well, which is not provided by the mode-dependent signal delay method.

Measurement of the input MD vector $\vec{\tau}_s(\omega)$ enables the determination of the DMGDs and the input PMs, exclusively, as shown in (8). In this respect, the input MD vector $\vec{\tau}_s(\omega)$ indeed encapsulates the modal dispersion of the fiber. In contrast, the fiber transfer matrix $\mathbf{U}(\omega)$ inherently contains additional information, e.g., one can also determine the output PMs, as well.

Based on the above discussion, we conclude that the mode-dependent signal delay method does not provide the same information as alternative methods measuring the fiber transfer matrix, e.g., swept wavelength interferometry [15]–[17].

F. Linking the Input MD Vector to the Pulse Group Delay

Following the methodology of [10] and [19], assuming perfectly coherent signals, we can prove that the group delay τ_g of an optical pulse with carrier angular frequency ω propagating through the optical fiber under test is related to the input MD vector $\vec{\tau}_s$ and the Stokes vector $\hat{s}$ representing the combination of launch modes [14]

$$\tau_g = \tau_0 + \frac{1}{2C_N^2}\langle\vec{\tau}_s\rangle \cdot \hat{s}, \tag{12}$$

where all quantities depend on the carrier angular frequency ω but we omitted this dependence for notational simplicity.

In (12), the group delay τ_g is defined as the first moment in time [10]

$$\tau_g := \frac{1}{\bar{E}} \int_{-\infty}^{\infty} t \mathbf{E}_0(t)^\dagger \mathbf{E}_0(t) \, dt, \tag{13}$$

where $\mathbf{E}_0(t)$ is the electric field at the fiber output and $\bar{E}$ is the average energy of the received pulse

$$\bar{E} := \int_{-\infty}^{\infty} \mathbf{E}_0(t)^\dagger \mathbf{E}_0(t) \, dt. \tag{14}$$

Finally, in (12), we defined the spectrally-averaged input MD vector $\langle \vec{\tau}_s \rangle$ as

$$\langle \vec{\tau}_s \rangle := \frac{1}{2\pi\bar{E}} \int_{-\infty}^{\infty} \vec{\tau}_s(\omega') |G(\omega')|^2 d\omega', \tag{15}$$

where $G(\omega)$ is the input pulse spectrum.

Our expression (12) differs from the initial expression (16) of Milione *et al.* [19] on two important points: the input MD vector $\vec{\tau}_s$ is spectrally-averaged and there is a corrective multiplicative factor of $1/2C_N^2$ in front of the inner product $\langle \vec{\tau}_s \rangle \cdot \hat{s}$.

In the following, we assume that we use optical pulses with sufficiently narrow spectrum around the carrier angular frequency ω so that $\langle \vec{\tau}_s \rangle \simeq \vec{\tau}_s$.

G. MD Vector Estimation

First, the average group delay τ_0 in (12) can be estimated by using the following procedure: We launch pulses corresponding to N arbitrary orthogonal states in Jones space $|s_{0,i}\rangle$ and measure their group delays $\tau_{0g,i}$, $i = 1, \ldots, N$. We know that N orthonormal vectors in Jones space are mapped into Stokes vectors $\hat{s}_{0,i}$ that form the vertices of a $(N-1)$-dimensional regular simplex [25]. This implies that $\hat{s}_{0,i}$ sum to zero [14]

$$\sum_{i=1}^{N} \hat{s}_{0,i} = \mathbf{0}. \tag{16}$$

Taking the average of the corresponding group delays $\tau_{0g,i}$ [14], expression (12) yields

$$\tau_0 = \frac{1}{N} \sum_{i=1}^{N} \tau_{0g,i}. \tag{17}$$

Subsequently, we launch $N^2 - 1$ linearly independent input states in Stokes space and measure the corresponding group delays τ_{g_i}, $i = 1, \ldots, N^2 - 1$. Expression (12) can be used to form a $(N^2 - 1) \times (N^2 - 1)$ system of linear equations. We can represent these equations in matrix form. First, we define the coefficient matrix

$$\mathbf{S} := [\hat{s}_1, \ldots, \hat{s}_{N^2-1}]^T. \tag{18}$$

Notice that the columns of $\mathbf{S}^T$ are the launch states represented by the Stokes vectors $\hat{s}_i$, $i = 1, \ldots, N^2 - 1$.

Then, we define the column vector of the DMGD's

$$\mathbf{T}_g := 2C_N^2 \left[\tau_{g,1} - \tau_0, \ldots, \tau_{g,N^2-1} - \tau_0 \right]^T. \tag{19}$$

Finally, the matrix representation of the system of equations is written as

$$\mathbf{S}\vec{\tau}_s = \mathbf{T}_g. \tag{20}$$

Thus, the MD vector $\vec{\tau}_s$ is given by

$$\vec{\tau}_s = \mathbf{S}^{-1}\mathbf{T}_g. \tag{21}$$

For $\mathbf{S}$ to be invertible, its determinant must be nonzero, $\det(\mathbf{S}) \neq 0$, so $\hat{s}_i$, $i = 1, \ldots, N^2 - 1$, have to be linearly independent.

H. Noise Modeling

Let us investigate the effect of the thermal noise $n(t)$ of the direct-detection receiver on the measurement of the DMGD's.

The error in the estimate of the group delay τ_g is

$$\delta\tau_g = \frac{1}{R_d\bar{E}} \int_{-T/2}^{T/2} t n(t) dt, \tag{22}$$

where R_d denotes the responsivity of the photodiode and T denotes the integration time for the computation of the group delay τ_g.

We assume that the thermal noise at the direct-detection receiver can be modeled as an additive white Gaussian noise (AWGN) with zero mean and autocorrelation function given by [26]

$$R(t, t') := E\{n(t)n(t')\} = \frac{N_0}{2}\delta(t - t'), \tag{23}$$

where the operator $E\{.\}$ denotes the expected value, $N_0/2$ denotes the power spectral density of the noise, and $\delta(t)$ is the Dirac delta function.

The mean of $\delta\tau_g$ is calculated by taking the expectation of both sides of (22) [26]

$$\mu_{\delta\tau_g} := E\{\delta\tau_g\} = \frac{1}{R_d\bar{E}} \int_{-T/2}^{T/2} t E\{n(t)\} dt = 0. \tag{24}$$

The variance of $\delta\tau_g$ is given by [26]

$$\sigma_{\delta\tau_g}^2 := \frac{1}{R_d^2\bar{E}^2} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} tt' E\{n(t)n(t')\} dtdt'. \tag{25}$$

Upon substituting (23) into (25), we find

$$\sigma_{\delta\tau_g}^2 = \frac{N_0}{2R_d^2\bar{E}^2} \int_{-T/2}^{T/2} t^2 dt = \frac{N_0 T^3}{24 R_d^2 \bar{E}^2}. \tag{26}$$

The presence of thermal noise at the individual measurements $\tau_{g,i}$, $i = 1, \ldots, N^2 - 1$, can lead to a random offset $\delta\mathbf{T}_g$ in the estimation of the DMGD matrix $\mathbf{T}_g$ in (19). Consequently, there is an error in the estimate of $\vec{\tau}_s(\omega)$ in (21), namely

$$\delta\vec{\tau}_s = \mathbf{S}^{-1}\delta\mathbf{T}_g = \mathbf{A}\delta\mathbf{T}_g, \tag{27}$$

where we set $\mathbf{A} := \mathbf{S}^{-1}$ for brevity.

Taking the expectation of both sides of (27) and substituting (24) in (27), we obtain the mean of $\delta\vec{\tau}_s$

$$\boldsymbol{\mu}_{\delta\vec{\tau}_s} := E\{\delta\vec{\tau}_s\} = \mathbf{A} E\{\delta\mathbf{T}_g\} = \mathbf{0}. \tag{28}$$

The covariance matrix of $\delta\vec{\tau}_s$ is given by

$$\mathbf{C}_{\delta\vec{\tau}_s} := \mathrm{E}\left\{\delta\vec{\tau}_s\delta\vec{\tau}_s^T\right\} = \mathbf{A}\mathbf{C}_{\delta\mathbf{T}_g}\mathbf{A}^T, \qquad (29)$$

where we defined the covariance of $\delta\mathbf{T}_g$

$$\mathbf{C}_{\delta\mathbf{T}_g} := \mathrm{E}\left\{\delta\mathbf{T}_g\delta\mathbf{T}_g^T\right\}. \qquad (30)$$

We assume that the measurements of the components T_{g_i} of the column vector $\mathbf{T}_g$ are performed sequentially. As a result, the noise realizations $n_i(t)$ in different measurements are independent and the random offsets δT_{g_i} are uncorrelated. Therefore, the covariance matrix of $\delta\mathbf{T}_g$ is diagonal with elements $\sigma_{\delta T_g}^2$

$$\mathbf{C}_{\delta\mathbf{T}_g} = \sigma_{\delta T_g}^2\mathbf{I}, \qquad (31)$$

where

$$\sigma_{\delta T_g}^2 = 4C_N^4\sigma_{\delta\tau_g}^2. \qquad (32)$$

In (32), we assumed, for simplicity, that τ_0 is a deterministic (i.e., a perfectly measured) quantity, as opposed to the group delays $\tau_{g_i}, i = 1, \ldots, N^2 - 1$, that are independent, identically-distributed random variables. The reason for this approximation is that the variance $\sigma_{\delta\tau_0}^2 = \sigma_{\delta\tau_g}^2/N$, so it is much smaller than $\sigma_{\delta\tau_g}^2$ for large values of N. For small N's, it might be necessary to repeat the process for measuring τ_0 a certain number of times k and average the results. Then, $\sigma_{\delta\tau_0}^2 = \sigma_{\delta\tau_g}^2/(kN)$, so that $\sigma_{\delta\tau_0}^2$ is negligible compared to $\sigma_{\delta\tau_g}^2$ for large values of kN.

From (29), (31), it follows immediately that

$$\mathbf{C}_{\delta\vec{\tau}_s} = \sigma_{\delta T_g}^2\mathbf{A}\mathbf{A}^T. \qquad (33)$$

The diagonal entries of the covariance matrix $\mathbf{C}_{\delta\vec{\tau}_s}$ are the variances of the components of the random vector $\delta\vec{\tau}_s$. Thus, the variance of $\|\delta\vec{\tau}_s\|$ is given by

$$\sigma_{\|\delta\vec{\tau}_s\|}^2 := E\left\{\|\delta\vec{\tau}_s\|^2\right\} = \mathrm{Tr}\left[\mathbf{C}_{\delta\vec{\tau}_s}\right], \qquad (34)$$

where the operator $\mathrm{Tr}(.)$ denotes the trace of a matrix.

Substituting (33), we obtain

$$\sigma_{\|\delta\vec{\tau}_s\|}^2 = \sigma_{\delta T_g}^2\,\mathrm{Tr}\left[\mathbf{A}\mathbf{A}^T\right], \qquad (35)$$

We can calculate explicitly $\mathrm{Tr}\left[\mathbf{A}\mathbf{A}^T\right]$ from the singular value decomposition (SVD) of $\mathbf{S}$ [27]

$$\mathbf{S} = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^T, \qquad (36)$$

where $\mathbf{W}, \mathbf{V}$ are orthogonal matrices and $\mathbf{\Sigma}$ is a diagonal matrix with entries the singular values $\sigma_i, i = 1, \ldots, N^2 - 1$, of $\mathbf{S}$.

Then, $\mathbf{A} = \mathbf{S}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}^T$ and we have

$$\mathbf{A}\mathbf{A}^T = \mathbf{V}\mathbf{\Sigma}^{-2}\mathbf{V}^T. \qquad (37)$$

Hence, we obtain a succinct expression for the trace

$$\mathrm{Tr}\left[\mathbf{A}\mathbf{A}^T\right] = \sum_{k=1}^{N^2-1}\sigma_k^{-2}. \qquad (38)$$

Combining (35) and (38) yields

$$\sigma_{\|\delta\vec{\tau}_s\|}^2 = \sigma_{\delta T_g}^2\sum_{k=1}^{N^2-1}\sigma_k^{-2}. \qquad (39)$$

To find a lower bound for (39), we use the fact that the arithmetic mean of the numbers σ_k^{-2} is greater than or equal to the geometric mean of the set (arithmetic/geometric mean inequality)

$$\frac{1}{N^2-1}\sum_{k=1}^{N^2-1}\sigma_k^{-2} \geq \left(\prod_{k=1}^{N^2-1}\sigma_k^{-2}\right)^{1/(N^2-1)}. \qquad (40)$$

In addition,

$$\prod_{k=1}^{N^2-1}\sigma_k^2 = \det(\mathbf{S}\mathbf{S}^T) = \det(\mathbf{S})^2 \leq 1, \qquad (41)$$

because $|\det(\mathbf{S})|$ is the volume of the N^2-1-dimensional parallelotope spanned by the unit Stokes vectors $\hat{s}_j$, which cannot exceed that of a cube.

Therefore,

$$\sum_{k=1}^{N^2-1}\sigma_k^{-2} \geq N^2 - 1. \qquad (42)$$

Notice that, if $\mathbf{S}$ is an orthogonal matrix, then so is $\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. Thus, we obtain the minimum estimated variance from (35)

$$\sigma_{\|\delta\vec{\tau}_s\|}^2 = (N^2 - 1)\sigma_{\delta T_g}^2. \qquad (43)$$

In addition, since the off-diagonal elements of $\mathbf{A}\mathbf{A}^T$ are zero in this case, the components of $\delta\vec{\tau}_s$ are uncorrelated and therefore independent.

In contrast, if $\mathbf{S}$ is not an orthogonal matrix, then $\sigma_{\|\delta\vec{\tau}_s\|}^2 > (N^2 - 1)\sigma_{\delta T_g}^2$, i.e., there is noise amplification compared to the previous case, which is due to the matrix inversion in (21). Moreover, the off-diagonal elements of $\mathbf{A}\mathbf{A}^T$ are nonzero and the components of $\delta\vec{\tau}_s$ become correlated.

I. Joint Measurement of MD and MDL

In our previous formulation, we assumed that MDL was negligible and that the fiber transfer matrix was unitary $\mathbf{U}(\omega)$. This was an instructive special case. In the presence of MDL, the fiber transfer matrix becomes non-unitary and is denoted by $\mathbf{H}(\omega)$. The output group delay operator then becomes $i\mathbf{H}_\omega(\omega)\mathbf{H}^{-1}(\omega)$ [28]. The corresponding input group delay operator becomes $i\mathbf{H}^{-1}(\omega)\mathbf{H}_\omega(\omega)$. The latter is a non-Hermitian matrix, so it has complex eigenvalues and non-orthogonal eigenvectors, in general [28].

Following closely the methodology of [28], we can represent the input group delay operator as a linear combination of the identity matrix and the Gell-Mann matrices

$$i\mathbf{H}^{-1}(\omega)\mathbf{H}_\omega(\omega) = \chi_0(\omega)\mathbf{I} + \frac{1}{2C_N}\vec{\chi}_s(\omega)\cdot\mathbf{\Lambda}, \qquad (44)$$

where $\chi_0(\omega)$, $\vec{\chi}_s(\omega)$ denote the complex mean group delay and the complex MD vector, respectively.

The input PMs are the eigenstates of the group-delay operator [28]

$$\frac{1}{2C_N}\vec{\chi}_s(\omega)\cdot\mathbf{\Lambda}\,|\,\chi_i(\omega)\rangle = \chi_i(\omega)|\,\chi_i(\omega)\rangle,\qquad(45)$$

The DMGDs are given by [28]

$$\tau_i(\omega) = \Re\left[\chi_i(\omega)\right],\ i = 1,\ldots,N.\qquad(46)$$

We can use the right polar decomposition of the fiber transfer matrix $\mathbf{H}(\omega)$ [28]

$$\mathbf{H}(\omega) = \mathbf{U}(\omega)\mathbf{P}(\omega),\qquad(47)$$

where $\mathbf{U}(\omega)$ is a unitary matrix and $\mathbf{P}(\omega)$ is a positive-semidefinite Hermitian matrix.

Multiplying $\mathbf{H}(\omega)$ by its adjoint yields

$$\mathbf{H}(\omega)^\dagger\mathbf{H}(\omega) = \mathbf{P}(\omega)^\dagger\mathbf{P}(\omega) = \mathbf{P}(\omega)^2.\qquad(48)$$

Since $\mathbf{P}(\omega)$ is a Hermitian positive-semidefinite matrix, its eigenvalues are real and its eigenvectors corresponding to different eigenvalues are orthogonal to each other.

Using the spectral decomposition of $\mathbf{P}(\omega)^2$ in terms of its eigenvalues $\exp\left[-a_k(\omega)z\right]$ and eigenvectors $|v_k(\omega)\rangle$, we obtain

$$\mathbf{P}(\omega)^2 = \sum_{k=1}^{N} e^{-a_k(\omega)z}|v_k(\omega)\rangle\langle v_k(\omega)\,|,\qquad(49)$$

where z represents the fiber length.

We refer to $a_k(\omega)$, $|v_k(\omega)\rangle$ as the *principal attenuation coefficients* and the *principal attenuation modes*, respectively. The latter are pairwise orthogonal.

We define MDL as the ratio of the largest eigenvalue to the smallest eigenvalue of $\mathbf{P}(\omega)^2$

$$\mathrm{MDL} := \frac{\max\left\{e^{-a_k(\omega)z}\right\}_{k=1}^{N}}{\min\left\{e^{-a_k(\omega)z}\right\}_{k=1}^{N}}.\qquad(50)$$

This is identical to taking the SVD of the fiber transfer matrix $\mathbf{H}(\omega)$ and defining MDL as the ratio of the squares of the largest and the smallest singular values [16].

From (4), the projection operators $|v_k(\omega)\rangle\langle v_k(\omega)\,|$ can be decomposed into the basis of the identity matrix and the Gell-Mann matrices [14]

$$|v_k(\omega)\rangle\langle v_k(\omega)\,| = \frac{1}{N}\mathbf{I} + \frac{1}{2C_N}\hat{v}_k(\omega)\cdot\mathbf{\Lambda}.\qquad(51)$$

Therefore, combining (49) with (51) yields [29]

$$\mathbf{P}(\omega)^2 = \alpha_0(\omega)\left[\mathbf{I} + \frac{1}{2C_N}\vec{\Gamma}(\omega)\cdot\mathbf{\Lambda}\right],\qquad(52)$$

where we defined the mean attenuation as the arithmetic mean of the eigenvalues of $\mathbf{P}(\omega)^2$

$$\alpha_0(\omega) := \frac{1}{N}\sum_{k=1}^{N} e^{-a_k(\omega)z},\qquad(53)$$

and the MDL vector

$$\vec{\Gamma}(\omega) := \frac{1}{\alpha_0(\omega)}\sum_{k=1}^{N} e^{-a_k(\omega)z}\hat{v}_k(\omega).\qquad(54)$$

We can determine α_0, $\vec{\Gamma}(\omega)$ by sending CW light into the fiber under test and measuring the average output power

$$\bar{P}(z) = \bar{P}(0)\langle s|\mathbf{H}(\omega)^\dagger\mathbf{H}(\omega)|s\rangle,\qquad(55)$$

or, equivalently, using (52), we can define the attenuation $\alpha(z)$ as

$$\alpha(z) := \frac{\bar{P}(z)}{\bar{P}(0)} = \alpha_0(\omega)\left[1 + \frac{1}{2C_N^2}\vec{\Gamma}(\omega)\cdot\hat{s}\right].\qquad(56)$$

Due to the similarity of (56) to (12), $\alpha_0(\omega)$, $\vec{\Gamma}(\omega)$ can be estimated using a similar process to the one described in Section II-G. Now, we simply measure the average output power instead of the time-of-flight of pulses. An alternative method for estimating MDL based on power measurements using a direct-detection receiver is described in [30].

Once $\alpha_0(\omega)$, $\vec{\Gamma}(\omega)$ are known, we can construct $\mathbf{H}(\omega)^\dagger\mathbf{H}(\omega)$ and calculate

$$\mathbf{P}(\omega) = \sqrt{\mathbf{H}(\omega)^\dagger\mathbf{H}(\omega)}.\qquad(57)$$

We can optically compensate for the optical fiber's MDL at the fiber input. The transfer matrix of the MDL-equalized optical fiber will then be

$$\mathbf{U}(\omega) = \mathbf{H}(\omega)\mathbf{P}(\omega)^{-1}.\qquad(58)$$

Since $\mathbf{U}(\omega)$ is unitary, the input group delay operator $i\mathbf{U}^\dagger(\omega)\mathbf{U}_\omega(\omega)$ can be expressed as a linear combination of the identity matrix and the Gell-Mann matrices as in (7)

$$i\mathbf{U}^\dagger(\omega)\mathbf{U}_\omega(\omega) = \tau_0'(\omega)\mathbf{I} + \frac{1}{2C_N}\vec{\tau}_s{}'(\omega)\cdot\mathbf{\Lambda}.\qquad(59)$$

Using the process described in Section II-G, we can estimate $\tau_0'(\omega)$, $\vec{\tau}_s{}'(\omega)$.

By differentiating (47) with respect to the angular frequency ω, we obtain

$$i\mathbf{H}(\omega)^{-1}\mathbf{H}_\omega(\omega) = \mathbf{P}^{-1}(\omega)\left[i\mathbf{U}^\dagger(\omega)\mathbf{U}_\omega(\omega)\right]\mathbf{P}(\omega)$$
$$+ i\mathbf{P}^{-1}(\omega)\mathbf{P}_\omega(\omega).$$

Since we have already calculated $\mathbf{P}(\omega)$, $i\mathbf{U}^\dagger(\omega)\mathbf{U}_\omega(\omega)$, we can calculate $i\mathbf{H}(\omega)^{-1}\mathbf{H}_\omega(\omega)$ and its eigenvalues and eigenvectors.

In summary, in the joint presence of MD and MDL, we follow a divide-and-conquer approach, i.e., we conduct the experiment into two successive phases: (i) Initially, we perform MDL characterization exclusively, followed by optical compensation of MDL; (ii) Then, we measure the MD vector of the compensated fiber transfer matrix. At each stage, we perform sequential measurements for the various components of the MDL and MD vectors at different frequencies, following identical procedures, launching the same set of quasi-orthogonal vectors each time. Finally, we combine the experimental results to create the input group delay operator and calculate the input PMs and the corresponding DMGDs of the optical fiber.

J. Practical Considerations

1) Modal Crosstalk: The mode-dependent signal delay method is vulnerable to errors in the mode converter settings. In this subsection, we aim to assess the impact of modal crosstalk induced by erroneous mode converter settings. A back-of-the-envelope calculation, for the case of SMFs ($N = 2$), indicates that this type of crosstalk can be the most significant limiting factor of the mode-dependent signal delay method. However, it can also be minimized by using quasi-orthogonal vectors in Stokes space as launch modes.

For MD characterization of FMFs and MMFs, mode selection at the fiber input can be accomplished, for instance, by using a Liquid Crystal on Silicon (LCOS)-based spatial light modulator (SLM) as a mode converter [31]. These devices can be used to implement reconfigurable phase masks by varying the voltage of their pixels. Their main drawback is that they exhibit high crosstalk between certain pairs of modes [31]. The optimization of phase masks using simulated annealing can reduce the crosstalk introduced by the SLM [17]. Nevertheless, any residual crosstalk in the transfer matrix of the mode converter may influence the accuracy of the measurement of the fiber input MD vector.

Detailed physical modeling of the operation of the SLM is out of the scope of this paper. We consider an abstract model instead. Assume that we launch the perturbed states $\hat{s}'_i$ instead of the intended states $\hat{s}_i$, due to errors in the SLM settings. Then, we obtain erroneous group delay measurements that affect the entries of the DMGD vector $\mathbf{T}_g$ in (19). The elements of the modified DMGD vector $\mathbf{T}'_g$ are given by

$$\mathbf{T}'_g = \mathbf{S}'\vec{\tau}_s, \tag{60}$$

where we defined the perturbed coefficient matrix

$$\mathbf{S}' := \left[\hat{s}'_1, \ldots, \hat{s}'_{N^2-1} \right]^T. \tag{61}$$

Assuming that we are unaware of the errors in the SLM settings, we reconstruct the MD vector using the matrix inversion (21)

$$\vec{\tau}_s' = \mathbf{S}^{-1} \mathbf{T}'_g.$$

The computed MD vector $\vec{\tau}_s'$ is offset from its nominal value $\vec{\tau}_s$ by

$$\delta\vec{\tau}_s = \vec{\tau}_s' - \vec{\tau}_s = \mathbf{S}^{-1}\delta\mathbf{S}\vec{\tau}_s, \tag{62}$$

where we defined the error matrix $\delta\mathbf{S}$

$$\delta\mathbf{S} := \mathbf{S}' - \mathbf{S}. \tag{63}$$

We want to estimate the error in (62). For this purpose, we will use the following inequality for square matrices [32, Th. 2.10]

$$\|\mathbf{AB}\| \le \|\mathbf{A}\| \, \|\mathbf{B}\|, \tag{64}$$

where $\|\mathbf{X}\|$ denotes the spectral norm of a matrix $\mathbf{X}$, which is the maximum singular value of $\mathbf{X}$.

Taking the spectral norm of both sides of (62) yields

$$\|\delta\vec{\tau}_s\| \le \|\mathbf{S}^{-1}\| \, \|\delta\mathbf{S}\| \, \|\vec{\tau}_s\|, \tag{65}$$

or, equivalently,

$$\frac{\|\delta\vec{\tau}_s\|}{\|\vec{\tau}_s\|} \le \|\mathbf{S}^{-1}\| \, \|\delta\mathbf{S}\| = \kappa\left(\mathbf{S}\right) \frac{\|\delta\mathbf{S}\|}{\|\mathbf{S}\|}, \tag{66}$$

where $\kappa\left(\mathbf{S}\right)$ is the condition number [32, Def. 2.12]

$$\kappa\left(\mathbf{S}\right) := \|\mathbf{S}^{-1}\| \, \|\mathbf{S}\|. \tag{67}$$

The condition number increases as the matrix gets closer to being singular. The choice of quasi-orthogonal Stokes vectors reduces $\kappa\left(\mathbf{S}\right)$ but could affect $\|\delta\mathbf{S}\|$ as well.

As an illustrative example, consider the elementary case of a SMF ($N = 2$). In this case, the mode converter can be substituted by a polarization controller and we can use the polarization-dependent signal delay method [18] for PMD characterization.

Assume that the intended launch vectors are linear horizontal, linear 45 deg, and right-circular SOPs, denoted by $|s_1\rangle = |e_x\rangle, |s_2\rangle = |e_{45°}\rangle, |s_3\rangle = |e_{RC}\rangle$, respectively.

Due to the finite extinction ratio of the polarization controller, every time we attempt to launch a given SOP, we excite also its orthogonal SOP. The polarization crosstalk level ε is defined as the ratio of the power launched at the orthogonal SOP and the nominal signal power. Since Jones vectors represent electric fields, the generated SOPs $|s'_i\rangle, i = 1, \ldots, 3$ contain perturbations of order $\sqrt{\varepsilon}$. Modeling the transfer matrix of the polarization controller as unitary, the actual launch vectors are

$$\left|s'_1\right\rangle = \sqrt{1-\varepsilon}|e_x\rangle + \sqrt{\varepsilon}\,|e_y\rangle$$

$$\left|s'_2\right\rangle = \sqrt{1-\varepsilon}|e_{45°}\rangle + \sqrt{\varepsilon}|e_{-45°}\rangle$$

$$\left|s'_3\right\rangle = \sqrt{1-\varepsilon}|e_{RC}\rangle + \sqrt{\varepsilon}|e_{LC}\rangle$$

where $|e_y\rangle, |e_{-45°}\rangle, |e_{LC}\rangle$, denote the linear vertical, linear -45 deg, and left-circular SOPs, respectively.

In the ideal case without polarization crosstalk, the coefficient matrix $\mathbf{S}$ given by (18) is just the identity matrix. In the presence of crosstalk, the actual coefficient matrix $\mathbf{S}'$ is

$$\mathbf{S}' = \begin{bmatrix} 1-2\varepsilon & 2\sqrt{(1-\varepsilon)\varepsilon} & 0 \\ 2\sqrt{(1-\varepsilon)\varepsilon} & 1-2\varepsilon & 0 \\ 2\sqrt{(1-\varepsilon)\varepsilon} & 0 & 1-2\varepsilon \end{bmatrix} \tag{68}$$

Upon substituting (68) into (63), we obtain

$$\delta\mathbf{S}' = \begin{bmatrix} -2\varepsilon & 2\sqrt{(1-\varepsilon)\varepsilon} & 0 \\ 2\sqrt{(1-\varepsilon)\varepsilon} & -2\varepsilon & 0 \\ 2\sqrt{(1-\varepsilon)\varepsilon} & 0 & -2\varepsilon \end{bmatrix}. \tag{69}$$

Expanding in Taylor series with respect to the crosstalk level ε yields

$$\delta\mathbf{S} = \begin{bmatrix} O\left(\varepsilon\right) & 2\sqrt{\varepsilon} + O\left(\varepsilon\right) & 0 \\ 2\sqrt{\varepsilon} + O\left(\varepsilon\right) & O\left(\varepsilon\right) & 0 \\ 2\sqrt{\varepsilon} + O\left(\varepsilon\right) & 0 & O\left(\varepsilon\right) \end{bmatrix}. \tag{70}$$

The spectral norm of the error matrix $\delta\mathbf{S}$ is

$$\|\delta\mathbf{S}\| = 2\sqrt{2\varepsilon} + O(\varepsilon). \tag{71}$$

The bound of the relative error in the MD vector is given by

$$\frac{\|\delta\vec{\tau}_s\|}{\|\vec{\tau}_s\|} \leq \|\mathbf{S}^{-1}\| \ \|\delta\mathbf{S}\| = 2\sqrt{2\varepsilon} + O(\varepsilon). \tag{72}$$

We observe that the relative error in the PMD vector is of order $\sqrt{\varepsilon}$. It is crucial that we know the launched SOPs precisely, that is with $\varepsilon < -40$ dB, in order for the relative error in the assessment of the PMD vector $\|\delta\vec{\tau}_s\|/\|\vec{\tau}_s\|$ to be of the order of 1%. For $N > 2$, modal crosstalk is more severe. This places extremely stringent requirements on the accuracy with which we know the launch states into the fiber. In practice, this means that the transfer matrix of the mode converter is an inseparable part of the optical fiber transfer matrix and influences the measurements. All MD characterization methods essentially measure the joint transfer matrix of the input/output spatial multiplexer/demultiplexer and the optical fiber [16].

2) Numerical Quadrature Accuracy: At the direct-detection receiver, the signal is sampled and the group delay τ_g is estimated by evaluating the integral (13) numerically. The computation can be carried out by using any numerical integration technique for equally-spaced subdivisions [33]. The accuracy of the integral depends on the sampling frequency and the vertical resolution of the real-time oscilloscope, as well as the pulse shape, the integration interval T, and the particular quadrature rule.

Assume that we launch ideal Gaussian pulses with half-width at the $1/e$ power point [24] equal to $T_0 = 10$ ns. The integration time in (22) is $T = 50$ ns. Using a low-end, real-time oscilloscope with 2 GHz bandwidth, 5 GSa/s sampling frequency, and 16 b vertical resolution, the relative error for detecting 0.1 ps group delays is less than 1% when using the composite 3/8 Simpson's rule for numerical quadrature [33]. Therefore, it is not necessary to employ a fast real-time oscilloscope or fs pulses in order to achieve sub-ps resolution.

3) Thermal Noise Significance: The reconstruction of the MD vector is influenced by the thermal noise of the individual measurements. For $N \sim 100$, $N \sim 10^4$ sequential measurements are required to recover the $N \sim 10^4$ components of the MD vector. During the MD vector reconstruction, the variances of the noises of individual measurements add up. The noise impact on the MD vector ends up being 40 dB higher than the impact of the noise on each individual measurement.

Consider a thermal-noise limited direct-detection receiver with noise-equivalent power [24] equal to 10 pW/$\sqrt{Hz}$. An integration time $T = 50$ ns corresponds to a digital filter with 100 MHz noise-equivalent bandwidth. We assume that the bandwidth of the photodiode is much larger than this value so that the thermal noise is essentially filtered digitally. For 10 mW received power, the rms noise per measurement given by (26) is on the order of 0.1 ps. Even for 100-mode SDM MMF, the rms noise given by (35) is on the order of 10 ps.

III. OPTIMIZATION FORMALISM

We seek to compute a set of $N^2 - 1$ Stokes vectors $\hat{s}_i$, $i = 1, \ldots, N^2 - 1$, that minimizes the variance of the MD vector.

A. Cost Function

Neglecting $\sigma^2_{\delta T_g}$ in (35), since it is dependent on the specific implementation of the direct-detection receiver, we adopt the squared Frobenius norm of $\mathbf{A}$ as a normalized cost function

$$\xi := \|\mathbf{A}\|^2_F = \mathrm{Tr}\left[\mathbf{A}\mathbf{A}^T\right]. \tag{73}$$

We want to minimize ξ subject to the constraint that the Stokes vectors in the matrix $\mathbf{S}$ must correspond to valid combinations of modes in the generalized Jones space.

An alternative form for the cost function (73) is

$$\xi = \mathrm{Tr}(\mathbf{G}^{-1}), \tag{74}$$

where we defined

$$\mathbf{G} := \mathbf{S}\mathbf{S}^T. \tag{75}$$

Notice that the entries of $\mathbf{G}$ are the inner products of the Stokes vectors, i.e., $\mathbf{G}_{jk} = \hat{s}_j \cdot \hat{s}_k$. In other words, $\mathbf{G}$ is the Gram matrix of the Stokes vectors $\hat{s}_j$.

Since $\xi \geq N^2 - 1$, where the lower bound of $\xi = N^2 - 1$ occurs in the ideal case of orthonormal vectors, we define the penalty for choosing a set of $N^2 - 1$ non-orthonormal Stokes vectors $\hat{s}_i$, $i = 1, \ldots, N^2 - 1$, as

$$\delta := \frac{\xi}{(N^2 - 1)}. \tag{76}$$

The latter quantity can be viewed as the noise amplification per degree of freedom.

B. Gradient Descent Method

Assume that the Jones vectors $|s_i\rangle$, $i = 1, \ldots, N^2 - 1$, and, thus, $\mathbf{S}$ and ξ, are functions of n real parameters $p_1, \ldots, p_n$. We can write the parameters in a concise form as a column vector

$$\mathbf{p} := [p_1, \ldots, p_n]^T. \tag{77}$$

We define the gradient of ξ as the column vector

$$\nabla\xi := \left[\frac{\partial\xi}{\partial p_1}, \ldots, \frac{\partial\xi}{\partial p_n}\right]^T. \tag{78}$$

The method of gradient descent [21] uses an iterative algorithm to calculate a minimum of the cost function ξ. Starting from a given point $\mathbf{p}^{(0)}$, it makes successive steps to points $\mathbf{p}^{(k)}$ by moving opposite to the direction of the gradient, until it reaches a local minimum:

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} - \mu^{(k)}\nabla\xi\left(\mathbf{p}^{(k)}\right), \tag{79}$$

where $\mu^{(k)}$ is a positive constant (adaptive step size) [21]. This iterative process is continued until the magnitude of the gradient falls below a certain threshold or until a maximum number of iterations is reached.

From (73), the components of the gradient of ξ can be written as

$$\frac{\partial \xi}{\partial p_r} = 2\mathrm{Tr}\left[\frac{\partial \mathbf{A}}{\partial p_r}\mathbf{A}^T\right]. \tag{80}$$

One can prove (80) by writing the trace expression as a scalar using index notation, taking the derivative with respect to p_r, and rewriting the final result in matrix form.

We also use the following matrix identity

$$\frac{\partial \mathbf{A}}{\partial p_r} = -\mathbf{A}\frac{\partial \mathbf{S}}{\partial p_r}\mathbf{A}. \tag{81}$$

The trace of a product of matrices is invariant under cyclic permutation of the matrices in the product [27]. Based on this property, we can rewrite (80) after substituting (81) as

$$\frac{\partial \xi}{\partial p_r} = -2\mathrm{Tr}\left[\frac{\partial \mathbf{S}}{\partial p_r}\mathbf{B}\right], \tag{82}$$

where we defined the auxiliary matrix

$$\mathbf{B} := \mathbf{A}\mathbf{A}^T\mathbf{A}. \tag{83}$$

Assume that only one Stokes vector of the set $\hat{s}_1, \ldots, \hat{s}_{N^2-1}$ is a function of the parameter p_r. Furthermore, assume that this is the k-th Stokes vector $\hat{s}_k$. After taking the trace in (82), only the product of the k-th row of the first matrix with the k-th column of the second matrix remains. We write in a shorthand manner

$$\frac{\partial \xi}{\partial p_r} = -2\frac{\partial \hat{s}_k^T}{\partial p_r}\mathbf{B}_k, \tag{84}$$

where $\mathbf{B}_k$ denotes the k-th column of the matrix $\mathbf{B}$.

In addition, we need to satisfy the constraint that the Stokes vectors making up the matrix $\mathbf{S}$ should correspond to valid combinations of modes in the generalized Jones space. We defer the discussion about how to take this constraint into account until Section IV.

C. Alternative Formulation Using Jones Vectors

To speed up numerical optimization, we find that it is computationally advantageous to express the elements of the gradient in terms of Jones vectors instead of Stokes vectors. This avoids unnecessary transitioning between Jones and Stokes spaces and eliminates the need for Gell-Mann matrices.

As a starting point, we first calculate the derivative of the Stokes vector $\hat{s}_k$ with respect to p_r by differentiating (2)

$$\frac{\partial \hat{s}_k}{\partial p_r} = 2C_N\Re\left\{\left\langle s_k|\boldsymbol{\Lambda}|\frac{\partial s_k}{\partial p_r}\right\rangle\right\}. \tag{85}$$

Furthermore, we notice that matrix $\mathbf{B}$ can be expressed as $\mathbf{B} = \mathbf{S}^T\mathbf{G}^{-2}$, so we rewrite (82) as

$$\frac{\partial \xi}{\partial p_r} = -2\mathrm{Tr}\left[\mathbf{G}^{-2}\frac{\partial \mathbf{S}}{\partial p_r}\mathbf{S}^T\right]. \tag{86}$$

In (86), we invoked the cyclical property of the trace and we pre-multiplied $\partial \mathbf{S}/\partial p_r$ with $\mathbf{G}^{-2}$ and post-multiplied with $\mathbf{S}^T$.

Assume that only the k-th Stokes vector $\hat{s}_k$ is a function of the parameter p_r. It follows that we can rewrite the previous expression into the form

$$\frac{\partial \xi}{\partial p_r} = -2\sum_j \mathbf{G}_{jk}^{-2}\left(\hat{s}_j \cdot \frac{\partial \hat{s}_k}{\partial p_r}\right). \tag{87}$$

We can rewrite the term inside the parenthesis in (87) using (85) and the property [A.13] in [10]

$$\hat{s}_j \cdot \frac{\partial \hat{s}_k}{\partial p_r} = 2C_N\Re\left\{\left\langle s_k|(\hat{s}_j \cdot \boldsymbol{\Lambda})|\frac{\partial s_k}{\partial p_r}\right\rangle\right\}. \tag{88}$$

Using (4), we obtain

$$\frac{\partial \xi}{\partial p_r} = -8C_N^2\Re\left[\sum_j \mathbf{G}_{jk}^{-2}\langle s_k|s_j\rangle\left\langle s_j|\frac{\partial s_k}{\partial p_r}\right\rangle\right]$$
$$+ \frac{8C_N^2}{N}\Re\left[\sum_j \mathbf{G}_{jk}^{-2}\left\langle s_k|\frac{\partial s_k}{\partial p_r}\right\rangle\right]. \tag{89}$$

Also recall that (6) gives

$$\mathbf{G}_{jk} = \hat{s}_j \cdot \hat{s}_k = 2C_N^2\left[|\langle s_j|s_k\rangle|^2 - \frac{1}{N}\right]. \tag{90}$$

From (89), (90), it is apparent that $\partial \xi/\partial p_r$ can be expressed in terms of Jones vectors exclusively. Furthermore, (89) uses the inverse of $\mathbf{G}$, which is symmetric and positive semi-definite, unlike (84) that is a function of $\mathbf{A}$, the inverse of $\mathbf{S}$. The advantage of (89) over (84) is that inversion of positive semi-definite matrices can be done via Cholesky decomposition [27], which is twice as fast as the LU decomposition used for general matrices. Finally, we shall see later on, when we discuss the projected gradient method, that due to the constraints $\langle s_j|s_j\rangle = 1$, the terms in (89) not only simplify slightly but the entire second sum in (89) can be omitted, since it is orthogonal to the constraint manifold (cf. Section IV).

IV. OPTIMIZATION ALGORITHMS

Our aim is to find an "almost orthogonal" matrix $\mathbf{S}$ that minimizes the cost function ξ. In other words, we want to compute a set of N^2-1 maximally-orthogonal Stokes vectors $\hat{s}_i$ that correspond to feasible combinations of propagating modes. This section is devoted to the description of two different gradient descent algorithms [21] that can accomplish this task.

A. Hyperspherical Coordinates and Unconstrained Gradient Descent

In the first algorithm, we parameterize the j-th unit Jones vector $|s_j\rangle$ by using $2N-2$ hyperspherical coordinates [14]

$$|s\rangle := \left[\cos(\phi_1), \sin(\phi_1)\cos(\phi_2)e^{i\theta_1}, \ldots,\right.$$
$$\left.\sin(\phi_1)\cdots\sin(\phi_{N-2})\sin(\phi_{N-1})e^{i\theta_{N-1}}\right]^T. \tag{91}$$

Furthermore, we define the parameter vector $\mathbf{p}$ that contains the coordinates ϕ_{jv} and θ_{jv} of all N^2-1 Stokes vectors. Then, we perform unconstrained optimization in a real space of $n = (N^2-1)\times(2N-2)$ dimensions using the method of gradient descent (79).

For large N's, when $\mathbf{p}^{(0)}$ is selected randomly, the matrix $\mathbf{S}$ might become almost singular, and the cost function can initially assume very high values. In this case, we find that the convergence of the gradient descent method can be accelerated by first using the normalized gradient and a constant step size μ in (79)

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} - \mu \frac{\nabla \xi \left(\mathbf{p}^{(k)} \right)}{\left\| \nabla \xi \left(\mathbf{p}^{(k)} \right) \right\|}, \qquad (92)$$

for $k = 0, 1, 2, \ldots$.

When the value of the cost function is decreased below a certain threshold, we revert back to (79), and the adaptive step $\mu^{(k)}$ is selected using the backtracking method [21].

B. Cartesian Coordinates and Projected Gradient Descent

In the second algorithm, we parametrize the j-th Jones vector $|s_j\rangle = (s_{jv})_{v=1}^N \in \mathbb{C}^N$ by the $2N$ real parameters $x_{jv} := \Re(s_{jv})$ and $y_{jv} := \Im(s_{jv})$. Therefore we have $n := (N^2 - 1) \times 2N$ parameters, which we arrange into one column vector $\mathbf{p} \in \mathbb{R}^n = \mathbb{R}^{2N} \times \cdots \times \mathbb{R}^{2N}$ by concatenating together the $N^2 - 1$ copies of $[x_{j1}, y_{j1}, \ldots, x_{jN}, y_{jN}]^T \in \mathbb{R}^{2N}$, $j = 1, \ldots, N^2 - 1$.

The modest price to pay in order to avoid using the numerically slower trigonometric functions in the parameterization, is the imposition of $N^2 - 1$ unit length constraints

$$\gamma_j(\mathbf{p}) := \langle s_j | s_j \rangle = \sum_{v=1}^{N} \left(x_{jv}^2 + y_{jv}^2 \right) = 1, \qquad (93)$$

where $j = 1, \ldots, N^2 - 1$.

This is to say that the parameter vector $\mathbf{p}$ is restricted to a $(N^2 - 1) \times (2N - 1)$ dimensional manifold $\mathcal{M}$ in $\mathbb{R}^n$ that is the Cartesian product of $N^2 - 1$ unit spheres in $\mathbb{R}^{2N}$. Any nonzero $\mathbf{p} \in \mathbb{R}^n$ can be projected into $\mathcal{M}$ by simply normalizing each block $[x_{j1}, y_{j1}, \ldots, x_{jN}, y_{jN}]^T$. We denote this projection by $\mathrm{proj}(\mathbf{p})$.

Furthermore, at every $\mathbf{p} \in \mathcal{M}$, we have $N^2 - 1$ vectors normal to $\mathcal{M}$ given by the unit normals to the individual spheres,

$$\mathbf{n}_j(\mathbf{p}) := [0, \ldots, 0, x_{j1}, y_{j1}, \ldots, x_{jN}, y_{jN}, 0, \ldots, 0]^T, \quad (94)$$

where $j = 1, \ldots, N^2 - 1$.

Alternatively, these vectors are the normalized gradients of individual constraints:

$$\mathbf{n}_j(\mathbf{p}) = \frac{1}{2} \nabla \gamma_j(\mathbf{p}) = \frac{1}{2} \left[\frac{\partial \gamma_j}{\partial p_r} \right]_{r=1}^n$$

$$= \left[\Re \left\langle s_j \left| \frac{\partial s_j}{\partial p_r} \right. \right\rangle \right]_{r=1}^n. \qquad (95)$$

Given any vector $\mathbf{v} \in \mathbb{R}^n$ attached at $\mathbf{p} \in \mathcal{M}$, we can readily decompose it into components that are tangent (parallel) and orthogonal (normal) to the constraint manifold $\mathcal{M}$, $\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp$, where

$$\mathbf{v}_\perp = \sum_{j=1}^{N^2-1} \left(\mathbf{n}_j(\mathbf{p})^T \mathbf{v} \right) \mathbf{n}_j(\mathbf{p}). \qquad (96)$$
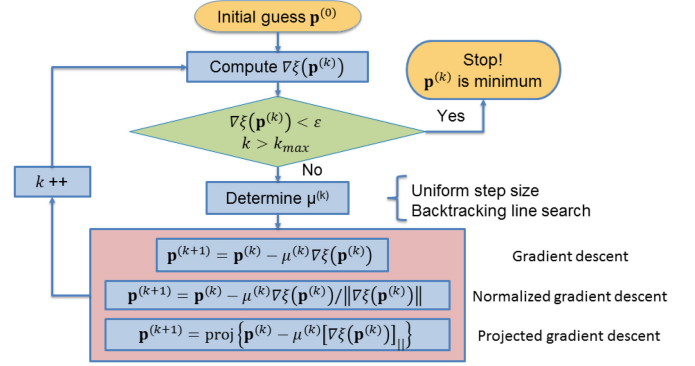

Fig. 2. Flowchart of the gradient descent methods used for computing quasi-orthonormal vector sets in Stokes space.

Recall that the gradient $\nabla \xi(\mathbf{p})$ is the column vector of all the partial derivatives given by (89) for $r = 1, \ldots, n$. In view of (95), the second sum in (89) is a linear combination of $\mathbf{n}_j(\mathbf{p})$. Thus it is orthogonal to $\mathcal{M}$ and can be omitted if we only need the tangential component of $\nabla \xi(\mathbf{p})$. Hence, to reduce the computational burden, we replace the gradient by

$$[\nabla \xi(\mathbf{p})]_\parallel = -8 C_N^2 \left[\Re \left\{ \sum_j \mathbf{G}_{jk}^{-2} \langle s_k | s_j \rangle \left\langle s_j \left| \frac{\partial s_k}{\partial p_r} \right. \right\rangle \right\} \right]_{r=1,\parallel}^n,$$

where the subscript $\parallel$ indicates taking the tangential component of the quantity inside the parenthesis.

The method of projected gradient descent starts from a given point $\mathbf{p}^{(0)} \in \mathcal{M}$ and then makes successive steps to points $\mathbf{p}^{(k)} \in \mathcal{M}$ by first moving in the direction opposite the tangential component of the gradient and then projecting (renormalizing) to hop back onto $\mathcal{M}$. The corresponding recursive formula is

$$\mathbf{p}^{(k+1)} = \mathrm{proj} \left\{ \mathbf{p}^{(k)} - \mu^{(k)} \left[\nabla \xi \left(\mathbf{p}^{(k)} \right) \right]_\parallel \right\} \qquad (97)$$

for $k = 0, 1, 2, \ldots$ [21].

To explore the sensitivity of both optimization algorithms to the choice of initial conditions, we perform several optimization runs with different random seeds. Given that the algorithms are computationally intensive, due to the high-dimensionality of the optimization space, individual runs are executed in parallel in a high-performance computing cluster.

The flowchart shown in Fig. 2 summarizes the steps of the numerical optimization process.

V. RESULTS AND DISCUSSION

In this section, we present the results of the numerical optimization algorithms described in Section IV.

As an illustrative example, the optimal launch mode combinations for $N = 4$ are given in Table I. Each mode is represented by a generalized Jones vector $|s_i\rangle, i = 1, \ldots, 15$, and its decomposition in terms of the fiber eigenmodes $|i\rangle, i = 1, \ldots, 4$, is listed. The angles among pairs of the corresponding generalized Stokes vectors vary in the interval $85°$–$97°$, so the generalized Stokes vectors are approximately orthogonal. The value of the

TABLE I
OPTIMAL VECTOR SET FOR $N = 4$

Vector set	Eigenmodes			
	$|1\rangle$	$|2\rangle$	$|3\rangle$	$|4\rangle$
$|s_1\rangle$	0.18	$0.52 - 0.5i$	$-0.07 + 0.53i$	$-0.4 - 0.03i$
$|s_2\rangle$	0.14	$0.39 + 0.57i$	$-0.06 - 0.57i$	$-0.04 + 0.42i$
$|s_3\rangle$	0.18	$-0.68 + 0.35i$	$0.02 + 0.43i$	$0.44 + 0.08i$
$|s_4\rangle$	0.23	$-0.51 - 0.57i$	$-0.01 - 0.45i$	$-0.01 - 0.38i$
$|s_5\rangle$	0.39	$-0.24 - 0.02i$	$-0.58 + 0.43i$	$0.04 - 0.51i$
$|s_6\rangle$	0.36	$-0.05 - 0.26i$	$0.51 - 0.51i$	$-0.52 + 0.04i$
$|s_7\rangle$	0.34	$0.11 - 0.02i$	$-0.64 + 0.47i$	$0.09 + 0.49i$
$|s_8\rangle$	0.34	$0.16i$	$0.52 - 0.59i$	$0.48 - 0.07i$
$|s_9\rangle$	0.45	$-0.09 + 0.32i$	$-0.04 - 0.32i$	$-0.41 - 0.64i$
$|s_{10}\rangle$	0.44	$-0.46 - 0.13i$	$-0.03 + 0.18i$	$-0.62 + 0.4i$
$|s_{11}\rangle$	0.42	$-0.06 - 0.41i$	$-0.04 - 0.16i$	$0.52 + 0.59i$
$|s_{12}\rangle$	0.42	$0.36 - 0.02i$	$-0.03 + 0.14i$	$0.6 - 0.55i$
$|s_{13}\rangle$	0.66	$-0.39 + 0.31i$	$-0.31 - 0.38i$	$0.23 + 0.13i$
$|s_{14}\rangle$	0.69	$-0.45 - 0.34i$	$0.29 + 0.24i$	$0.16 - 0.2i$
$|s_{15}\rangle$	0.69	$0.18 - 0.46i$	$-0.41 - 0.29i$	$-0.12 - 0.12i$

Note: Jones vectors $|s_i\rangle$, $i = 1, \ldots, 15$, corresponding to the optimal vector set for N=4. In each column we list their weights as a function of the fiber eigenmodes $|i\rangle$, $i = 1, \ldots, 4$.
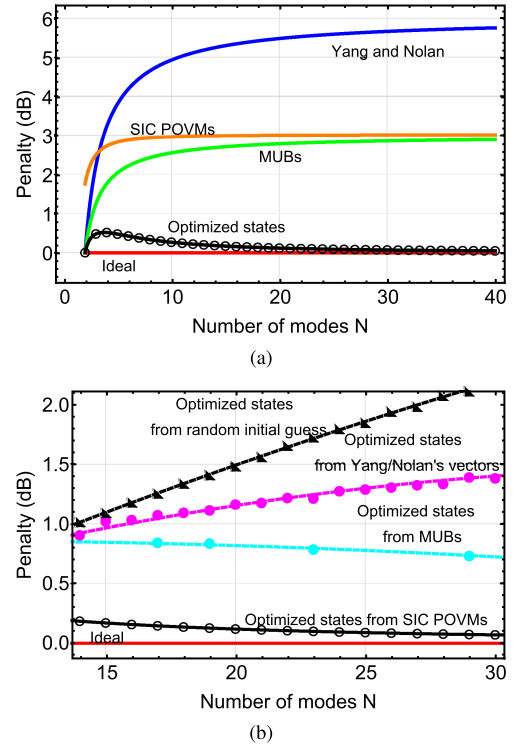

Fig. 3. (a) SNR penalty compared to the ideal case vs the number of modes for four different vector sets (Symbols: Blue line: Yang and Nolan's vectors [20]; Green line: MUBs [23]; Orange line: SIC POVMs [22]; Red line: Orthonormal Stokes vectors; Circles: Numerical optimization using the algorithms in Section IV. The optimization results were also validated by using the function FindMinimum in Mathematica [34], which is based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm. Early results were presented in [35]). (b) Sensitivity of the numerical optimization results to the initial conditions for $N = 14$–30 (Symbols for different initial conditions: Triangles: Random initial guess; Magenta points: Yang and Nolan's vectors; Cyan points: MUBs; Open circles: SIC POVMs).

cost function for this vector set is $\xi = 16.9$, while for a truly orthogonal vector set it would be $\xi = 15$. The SNR penalty is 0.517 dB. Penalties for the optimum vector sets for other values of N are given in Fig. 3(a).

Before we proceed, it would be instructive to explain how to use the results of Table I. Consider the case of an FMF supporting the LP$_{01}$ and LP$_{11}$ mode groups. If there is weak coupling between the FMF mode groups, one can choose to characterize the modal dispersion of the LP$_{01}$ and LP$_{11}$ mode groups separately. In this case, the mode-dependent signal delay method can be used first to determine the MD vector of the LP$_{01}$ mode group. This requires launching three different combinations of the x- and y-polarizations of the LP$_{01}$ mode. For this purpose, we can select three arbitrary orthonormal vectors in the conventional 3D Stokes space, e.g., the linear horizontal, linear 45 deg, and right-circular SOPs. Then, the mode-dependent signal delay method can be applied once more to determine the MD vector of the LP$_{11}$ mode group. It is well known that the LP$_{11}$ mode group is composed of four spatial and polarization modes, i.e., the LP$_{11,o}$ and LP$_{11,e}$ modes, each in two orthogonal polarization configurations. The mode-dependent signal delay method for $N = 4$ requires launching 15 launch mode combinations. Launching the 15 launch mode combinations $|s_i\rangle$, $i = 1, \ldots, 15$, shown in Table I will yield the smallest possible error in the measurement of the MD vector. The fiber eigenmodes $|i\rangle$, $i = 1, \ldots, 4$, are the constituents of the LP$_{11}$ mode group. The complex coefficients in Table I represent the complex excitations of the phasors of the electric fields of these modes.

Next, it is shown that the computed optimal Stokes vector sets yield much better performance than previously proposed vector sets. Fig. 3(a) shows plots of the SNR penalty δ as a function of the number of propagation modes N in the optical fiber for various vector sets. The ideal, albeit infeasible, case of orthonormal vectors is shown by the horizontal red line. The results of the numerical optimization are represented by the black curve with circles. Notice that the penalty is initially 0 dB for $N = 2$, reaches a maximum value for $N = 4$, and then falls monotonically to almost 0 dB for $N = 40$. The fact that

the penalty is 0 dB for $N = 2$ comes as no surprise: in this case, the whole surface of the Poincaré sphere is covered with valid states. Thus, there exists an infinity of orthonormal vector sets that can be used for the measurement of the MD vector in Stokes space. For larger values of N, it is impossible to find an orthonormal set of $N^2 - 1$ Stokes vectors. For instance, for $N = 4$, we observe that there is 0.517 dB penalty with respect to the ideal case. By further increasing N, we observe a gradual reduction in penalty, reaching 0.046 dB for $N = 40$.

For comparison, we included in the same graph, three additional plots corresponding to vector sets proposed in prior literature in optical communications and quantum mechanics, namely Yang and Nolan's vectors [20], vectors selected from MUBs [23], and SIC-POVM vectors [22] in blue, green, and orange, respectively (see Appendix A for details). The main advantage of these three vector sets is that there are relatively simple analytical or numerical algorithms for the evaluation of their coordinates. In contrast, computing the optimal vector sets using the method of gradient descent is time consuming for large values of N. On the downside, Yang and Nolan's vectors, MUBs, and SIC-POVMs present much higher penalties than the optimal vector sets given by numerical optimization. Indicatively, we remark the following features: a) $N^2 - 1$ Stokes

vectors from MUBs can be evaluated using various numerical algorithms only for values of N that are prime numbers or powers of prime numbers [23]. It is worth noting that this vector set can be used only for waveguides lacking cylindrical symmetry. For the optical fibers of interest, the number of modes N can never be a prime or a power of a prime. We refer to MUBs here just for completeness. The green line in Fig. 3(a) shows that there is an asymptotic penalty equal to 3 dB for large N's, i.e., the noise in the estimate of the MD vector is amplified by a factor of 2. b) SIC POVMs exhibit slightly worse performance with respect to MUB vectors for small values of N but, as N increases, the penalty asymptotically reaches a ceiling of 3 dB, as in the case of MUB vectors (orange line in Fig. 3(a)). The main advantage of this vector family, compared to MUBs, is that $N^2 - 1$ vector sets can be computed for all practical values of N. c) Yang and Nolan's vectors are given by simple analytical formulas, in contrast to the previous vector families. Nevertheless, they present worse performance than all prior vector sets, and the corresponding penalty asymptotically reaches 6 dB (blue line in Fig. 3(a)).

In summary, the optimal vector sets provided by numerical optimization increase the SNR of the measurements asymptotically by 3 dB for large values of N compared to SIC POVMs and MUBs and by about 6 dB compared to Yang and Nolan's vectors. Therefore, we conclude that the performance of the mode-dependent signal delay can be dramatically improved by using the optimal vector sets provided by the gradient descent method. One can contrast this finding with the claim in Yang and Nolan's paper [20] that the results of the mode-dependent signal delay method do not depend on the choice of launch state vectors, as long as the latter are linearly independent. This is true only in the absence of receiver noise.

Despite their inferior performance compared to the optimal vectors, MUBs, SIC POVMs, and Yang and Nolan's vectors are still useful as they can be used as starting points in order to accelerate the convergence of the gradient descent method. Fig. 3(b) compares the results of the gradient descent method when the aforementioned vector sets are used as initial guesses for $N = 15 - 30$. We observe that the gradient descent method converges to different local minima for each vector family after 100,000 iterations. Interestingly, using SIC POVMs as initial guesses leads to the lowest penalties (black curve with circles) whereas the use of random initial vectors as starting points leads to the worst performance (black curve with triangles). The use of MUB vectors and Yang and Nolan's vectors as starting points for the numerical optimization leads to intermediate penalty values in between the two back curves. Indicative values of penalties are shown in Table II.

In retrospect, it is not surprising that the gradient descent method yields best results when the SIC POVM vectors are used as an initial guess: This must be attributed to the fact that SIC POVMs present maximum symmetry because they form a regular simplex in Stokes space and are equiangular, i.e., their pairwise inner products in Stokes space are the same. As explained in detail in Appendix A, the pairwise inner product of two different SIC POVMs tends to zero for large values of N. This is illustrated by the almost diagonal Gram matrix

TABLE II
NUMERICAL OPTIMIZATION RESULTS (FIG. 3(B), $N = 30$)

Initial guess	Initial penalty (dB)	Final penalty (dB)
SIC POVMs	3.00	0.07
MUBs	2.86	0.73
Yang & Nolan's vectors	5.65	1.38
Random vectors	>50	2.2

Note: Starting from different initial conditions, the numerical optimization reaches different local minima after 100, 000 iterations. For instance, for $N = 30$, using SIC POVMs in the mode-dependent signal delay method leads to a 3 dB penalty compared to the ideal case. However, using the gradient descent method with the SIC POVMs as an initial guess, we compute an optimal set of vectors that exhibits only 0.07 dB residual penalty at the end of the optimization process compared to the ideal case. Worse residual penalties are achieved by starting the optimization process using vectors from MUBs, Yang & Nolan's vectors, and random vectors as initial conditions.

for $N = 5$ in Fig. 4(a). For comparison, density plots of the Gram matrices for vectors from MUBs and Yang and Nolan's vectors are shown in Fig. 4(b) and (c), respectively. The latter two density plots reveal a block diagonal and a block structure, respectively, indicative of much less symmetric vector configurations.

Finally, we can catch a glimpse of the optimum set of Stokes vectors using a 2D projection (e.g., see Fig. 5 for $N = 3$). We know that the $N^2 - 1$ Stokes vectors should be ideally orthonormal. From the optimal vectors given by the numerical optimization procedure, we compute a set of orthonormal vectors best approximating the optimal vectors. It is possible to project these orthonormal vectors onto a plane so that their projections have equal angular separations (dashed black vectors). Now we can superimpose on the same plane the projections of the actual optimal vectors given by the numerical optimization procedure (red vectors), as well as the projection of the manifold of allowed states on the surface of the Poincaré sphere (light green area) [5]. All vectors are bounded by the projection of the Poincaré sphere onto the plane (pink circular disk with unit radius).

VI. SUMMARY

In this article, we revised the mode-dependent signal delay method formalism for the characterization of SDM MMFs. We analytically calculated the variance in the estimation of the length of the input MD vector due to receiver thermal noise. We showed that the mode-dependent signal delay method is versatile and can be applied to the estimation of the MDL vector, as well as the simultaneous measurement of the MD and MDL vectors. We discussed various measurement errors other than these due to the thermal noise of the direct-detection receiver that occur during the characterization process. The latter part of the paper was devoted to the optimization of the launch states used in the mode-dependent signal delay method for the measurement of modal dispersion in SDM MMFs. The optimal sets of launch modes proposed here are universal, i.e., they are not limited to specific fiber types and can be used for SMFs, MMFs, and MCFs with strong and weak coupling.
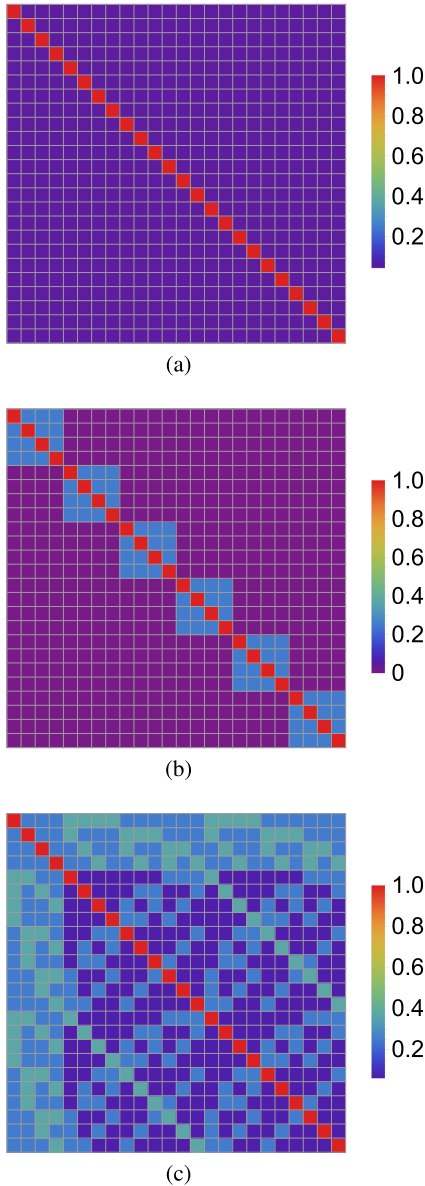
(a)

(b)

(c)

Fig. 4. Density plots of the absolute value of the Gram matrix $\mathbf{G}$ of various vector sets for $N = 5$. (a) SIC-POVMs [22]. (b) MUBs [23]. (c) Yang and Nolan's vectors [20].
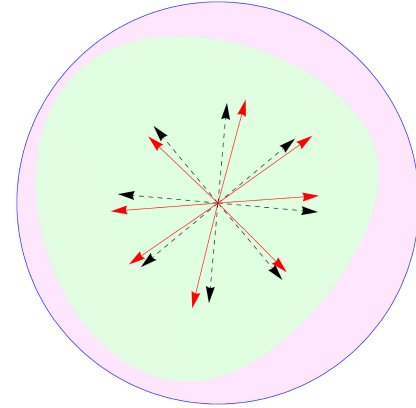

Fig. 5. 2D projections of various vector sets for $N = 3$ (Symbols: Red vectors: actual optimal vectors given by the numerical optimization of Section II; Dashed black vectors: ideal orthonormal vectors best approximating the optimal vectors; Light green area: projection of the manifold of allowed states on the surface of the Poincaré sphere).

APPENDIX A
SPECIAL VECTOR SETS

The goal of this Appendix is to derive useful analytical relationships for the cost function of special vector sets.

A. Symmetric, Informationally Complete, Positive Operator Valued Measure (SIC-POVM) Vectors [22]

Consider the N-dimensional Jones space $\mathbb{C}^N$. The inner product of the N^2 unit SIC-POVM vectors $\{|\psi_i\rangle\}$ satisfies the condition

$$|\langle\psi_i|\psi_j\rangle|^2 = \frac{1}{N+1}, \quad \forall i \neq j \tag{98}$$

We recall that the dot product of Stokes vectors is related to the inner product of Jones vectors through (6). By substituting (98) into (6), we can compute the exact dot products of the SIC POVM vectors in generalized Stokes space

$$\mathbf{G}_{ij} = \hat{\psi}_i \cdot \hat{\psi}_j = \begin{cases} 1 & i = j \\ -\dfrac{1}{N^2 - 1} & i \neq j \end{cases} \tag{99}$$

A density plot of the absolute value of the Gram matrix $\mathbf{G} = \mathbf{S}\mathbf{S}^T$ for $N = 5$ is shown in Fig. 4(a). Due to its simple structure, the cost function can be calculated analytically.

Namely, we can decompose $\mathbf{G}$ into a linear combination of two square $\left(N^2 - 1\right) \times \left(N^2 - 1\right)$ matrices, the identity matrix $\mathbf{I}$ and the constant matrix $\mathbf{J}$ with all entries equal to unity

$$\mathbf{G} = \frac{N^2}{N^2 - 1}\mathbf{I} - \frac{1}{N^2 - 1}\mathbf{J} \tag{100}$$

Matrices $\mathbf{I}$ and $\mathbf{J}$ commute so the eigenvalues of $\mathbf{G}$ are the corresponding combinations of the eigenvalues of these two matrices.

The identity matrix $\mathbf{I}$ has a unit eigenvalue with multiplicity $N^2 - 1$, and the matrix $\mathbf{J}$ has two eigenvalues, zero with multiplicity $N^2 - 2$ and $N^2 - 1$ with multiplicity 1. Therefore, the

As a final note, we stress that all MD characterization methods inherently make measurement errors due to receiver noise, modal crosstalk introduced during mode launch, and other implementation imperfections. Therefore, it is important to measure the modal dispersion of a fiber using various alternative methods and determine whether their results agree or not. Even if the mode-dependent signal delay method does not prevail as the method of choice for MD characterization, the set of optimal vectors proposed here can be used for measurements in the generalized Stokes space and data transmission using Stokes vector modulation. We anticipate that, since these vectors are quasi-orthonormal, they will give superior performance in a variety of problems involving measurements in Stokes space compared to other vector sets proposed in the literature.

Optimum vector sets are available online [36].

eigenvalues of $\mathbf{G}$ are

$$\lambda\left(\mathbf{G}\right) = \left\{\underbrace{\frac{N^2}{N^2-1}, \cdots, \frac{N^2}{N^2-1}}_{N^2-2}, \frac{1}{N^2-1}\right\}. \qquad (101)$$

Consequently, the eigenvalues of $\mathbf{G}^{-1}$ are

$$\lambda\left(\mathbf{G}^{-1}\right) = \left\{\underbrace{\frac{N^2-1}{N^2}, \ldots, \frac{N^2-1}{N^2}}_{N^2-2}, N^2-1\right\}. \qquad (102)$$

If $\mathbf{X}$ is a square $n \times n$ matrix, then the sum of the n eigenvalues of $\mathbf{X}$ is the trace of $\mathbf{X}$ and the product of the n eigenvalues is the determinant of $\mathbf{X}$.

Therefore, the cost function can be analytically expressed as

$$\xi = \mathrm{Tr}\left(\mathbf{G}^{-1}\right) = 2\frac{\left(N^2-1\right)^2}{N^2}. \qquad (103)$$

The penalty is given by

$$\delta = \frac{\xi}{N^2-1} = 2\frac{\left(N^2-1\right)}{N^2}. \qquad (104)$$

For optical fibers supporting a large number of modes N, the penalty asymptotically reaches the limit

$$\lim_{N\to\infty} \delta = 2. \qquad (105)$$

We conclude that, for large N's, there is roughly a 3 dB penalty compared to the ideal orthonormal states.

In addition, the volume of the parallelotope with edges equal to the Stokes vectors $\hat{s}_1, \ldots, \hat{s}_{N^2-1}$ is given by the determinant of the Gram matrix $V = \sqrt{\det\left(\mathbf{G}\right)}$. The last expression can be analytically evaluated from the eigenvalues in (102)

$$V = \sqrt{\det\left(\mathbf{G}\right)} = \frac{N^{N^2-2}}{\left(N^2-1\right)^{\frac{\left(N^2-1\right)}{2}}} \qquad (106)$$

For optical fibers supporting a large number of modes N, the volume of the parallelotope asymptotically tends to zero

$$\lim_{N\to\infty} V = 0. \qquad (107)$$

B. Vectors From Mutually Unbiased Bases (MUBs) [23]

Two distinct bases $\{|\psi_i\rangle\}, \{|\phi_j\rangle\}$ are said to be mutually unbiased if $\hat{\psi}_i \cdot \hat{\phi}_j = 0$. Then, from (6), we obtain

$$|\langle\psi_i|\phi_j\rangle|^2 = \frac{1}{N}, \quad \forall i, j. \qquad (108)$$

There exist $N+1$ MUBs of N vectors each when the number of modes N is a power of a prime [23]. Here, we select launch states by picking groups of $N-1$ vectors from each one of the $N+1$ MUBs.

From (6) and (108), we can calculate the elements of the covariance matrix $\mathbf{G} = \mathbf{SS}^T$ without first calculating explicitly the MUB vectors.

It turns out that $\mathbf{G} = \mathbf{SS}^T$ is a $\left(N^2-1\right) \times \left(N^2-1\right)$ square matrix in block diagonal form

$$\mathbf{G} = \begin{bmatrix} \mathbf{X} & 0 & 0 & 0 \\ 0 & \mathbf{X} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{X} \end{bmatrix}, \qquad (109)$$

where $\mathbf{X}$ is a $(N-1) \times (N-1)$ submatrix given by

$$\mathbf{X} = \begin{bmatrix} 1 & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \\ -\frac{1}{N-1} & 1 & \ddots & -\frac{1}{N-1} \\ \vdots & \ddots & \ddots & -\frac{1}{N-1} \\ -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} & 1 \end{bmatrix}. \qquad (110)$$

This form is analogous to (99). Therefore, we can decompose $\mathbf{X}$ as in (100)

$$\mathbf{X} = \frac{N}{N-1}\mathbf{I} - \frac{1}{N-1}\mathbf{J}, \qquad (111)$$

where now $\mathbf{I}, \mathbf{J}$ are square $(N-1) \times (N-1)$ matrices.

Following the same methodology as in the preceding subsection, it is straightforward to show that the eigenvalues of $\mathbf{X}$ are

$$\lambda\left(\mathbf{X}\right) = \left\{\underbrace{\frac{N}{N-1}, \ldots, \frac{N}{N-1}}_{N-2}, \frac{1}{N-1}\right\}. \qquad (112)$$

For a block diagonal matrix

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Y}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Y}_n \end{bmatrix}, \qquad (113)$$

the following properties hold

$$\det \mathbf{Y} = \prod_{i=1}^{n} \det \mathbf{Y}_i,$$

$$\mathrm{Tr}\mathbf{Y} = \sum_{i=1}^{n} \mathrm{Tr}\mathbf{Y}_i. \qquad (114)$$

Using (108)–(114), the cost function can be analytically expressed as

$$\xi = \mathrm{Tr}\left(\mathbf{G}^{-1}\right) = 2\left(N^2-1\right)\frac{N-1}{N}. \qquad (115)$$

The penalty is given by

$$\delta = \frac{\xi}{N^2-1} = 2\frac{N-1}{N}. \qquad (116)$$

For optical fibers supporting a large number of modes

$$\lim_{N \to \infty} \delta = 2 \tag{117}$$

Asymptotically, there is roughly a 3 dB penalty compared to the ideal orthonormal states.

In addition, the volume can be analytically expressed as

$$V = \sqrt{\det(\mathbf{G})} = \frac{N^{\frac{(N-2)(N+1)}{2}}}{(N-1)^{\frac{(N^2-1)}{2}}}. \tag{118}$$

For optical fibers supporting a large number of modes N, the volume of the parallelotope asymptotically tends to zero

$$\lim_{N \to \infty} V = 0. \tag{119}$$

C. Cost Function for Yang and Nolan's Vectors [20]

Consider the fiber eigenmodes in Jones space $|i\rangle$, $i = 1, \dots, N$. Yang and Nolan's vectors are defined as [20]

$$|x_i\rangle = |i\rangle \quad i = 1, \dots, N-1$$

$$|y_{ij}\rangle = \frac{|i\rangle + |j\rangle}{2} \quad 1 \le i < j \le N$$

$$|z_{ij}\rangle = \frac{|i\rangle + \mathrm{i}|j\rangle}{2} \quad 1 \le i < j \le N \tag{120}$$

In the last expression, we used different fonts to distinguish the imaginary number i from the index i. Notice that there are $N^2 - 1$ vectors in total.

The squared norms of the inner products in Jones space are

$$|\langle x_i \mid x_j \rangle|^2 = \delta_{ij}$$

$$|\langle x_i \mid y_{jk} \rangle|^2 = \frac{(\delta_{ij} + \delta_{ik})^2}{2}$$

$$|\langle x_i \mid z_{jk} \rangle|^2 = \frac{\delta_{ij} + \delta_{ik}}{2}$$

$$|\langle y_{ij} \mid y_{k\ell} \rangle|^2 = \frac{(\delta_{ik} + \delta_{i\ell} + \delta_{jk} + \delta_{j\ell})^2}{4}$$

$$|\langle y_{ij} \mid z_{k\ell} \rangle|^2 = \frac{(\delta_{ik} + \delta_{jk})^2 + (\delta_{i\ell} + \delta_{j\ell})^2}{4}$$

$$|\langle z_{ij} \mid z_{k\ell} \rangle|^2 = \frac{(\delta_{ik} + \delta_{j\ell})^2 + (\delta_{i\ell} - \delta_{jk})^2}{4} \tag{121}$$

The elements of the Gram matrix are given by (6)

$$\mathbf{G}_{jk} = \hat{s}_j \cdot \hat{s}_k = 2C_N^2 \left[|\langle s_j | s_k \rangle|^2 - \frac{1}{N} \right]. \tag{122}$$

Let's define the submatrices of Stokes vectors

$$X = [\hat{x}_1, \dots, \hat{x}_{N-1}]^T$$

$$Y = \left[\hat{y}_1, \dots, \hat{y}_{\frac{N(N-1)}{2}} \right]^T$$

$$Z = \left[\hat{z}_1, \dots, \hat{z}_{\frac{N(N-1)}{2}} \right]^T \tag{123}$$

where we reindexed the Stokes vectors from $\hat{y}_{ij}, \hat{z}_{ij}$ to $\hat{y}_k, \hat{z}_k$.

The Gram matrix structure is shown in Fig. 6. The Gram matrix is partitioned into nine blocks (in color). An analytical

Fig. 6. Partitioning of the Gram matrix for Yang and Nolan's vectors [20]. Rectangular submatrices are shown in brown and green, while all other colors indicate square submatrices of different dimensions.

calculation of the cost function in a way analogous to the case of SIC-POVMs and MUBs is cumbersome due to the more complex block structure of the Gram matrix. Therefore, the cost function shown in blue in Fig. 3(a) is calculated numerically by taking the trace of the inverted Gram matrix obtained by (121)–(123).

REFERENCES

[1] P. J. Winzer and D. T. Neilson, "From scaling disparities to integrated parallelism: A decathlon for a decade," *J. Lightw. Technol.*, vol. 35, no. 5, pp. 1099–1115, Mar. 2017.
[2] Cisco Systems Inc., "Cisco Visual Networking Index: Forecast and Methodology, 2016-2021," White Paper, Sep. 2017.
[3] A. Chraplyvy, "The coming capacity crunch," in *Proc. Eur. Conf. Opt. Commun.*, Vienna, Austria, 2009.
[4] P. J. Winzer, R. Ryf, and S. Randel, "Spatial multiplexing using multiple-input multiple-output signal processing," in *Optical Fiber Telecommunications VIB*. New York, NY, USA: Academic, 2013, ch. 10.
[5] T. Mizuno, H. Takara, K. Shibahara, A. Sano, and Y. Miyamoto, "Dense space division multiplexed transmission over multicore and multimode fiber for long-haul transport systems," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1484–1493, Mar. 2016.
[6] S. Mumtaz, R. J. Essiambre, and G. P. Agrawal, "Nonlinear propagation in multimode and multicore fibers: Generalization of the Manakov equations," *J. Lightw. Technol.*, vol. 31, no. 3, pp. 398–406, Feb. 2013.
[7] C. Antonelli, M. Shtaif, and A. Mecozzi, "Modeling of nonlinear propagation in space-division multiplexed fiber-optic transmission," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 36–54, Jan. 2016.
[8] R. Ryf *et al.*, "Long-haul transmission over multi-core fibers with coupled cores," in *Proc. Eur. Conf. Opt. Commun.*, Gothenburg, Sweden, 2017, pp. 1–3.
[9] C. D. Poole and J. Nagel, "Polarization effects in lightwave systems," in *Optical Fiber Telecommunications IIIA*. New York, NY, USA: Academic, 1997, ch. 6, pp. 114–161.
[10] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proc. Nat. Acad. Sci. USA*, 2000, vol. 97, pp. 4541–4550.

[11] H. Kogelnik, L. E. Nelson, and R. M. Jopson, "Polarization mode dispersion," in *Optical Fiber Telecommunications IV-B*, I. P. Kaminow and T. Li, Eds. New York, NY, USA: Academic, 2002.

[12] K.-P. Ho and J. M. Kahn, "Mode coupling and its impact on spatially multiplexed systems," in *Optical Fiber Telecommunications VI*. New York, NY, USA: Academic, 2013, ch. 11, pp. 491–568.

[13] C. Antonelli, A. Mecozzi, M. Shtaif, and P. J. Winzer, "Stokes-space analysis of modal dispersion in fibers with multiple mode transmission," *Opt. Express*, vol. 20, pp. 11718–11733, 2012.

[14] I. Roudas and J. Kwapisz, "Stokes space representation of modal dispersion," *IEEE Photon. J.*, vol. 9, no. 5, Oct. 2017, Art. no. 7203715.

[15] R. Ryf and N. K. Fontaine, "Space-division multiplexing and MIMO processing," in *Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks*, X. Zhou, and C. Xie, Eds. New York, NY, USA: Wiley, 2016, pp. 547–608.

[16] N. K. Fontaine *et al.*, "Characterization of space-division multiplexing systems using a swept-wavelength interferometer," in *Proc. Opt. Fiber Commun. Conf.*, Anaheim, CA, USA, 2013, Paper OW1K.2.

[17] J. Carpenter, B. J. Eggleton, and J. Schroder, "Observation of Eisenbud-Wigner-Smith states as principal modes in multimode fibre," *Nature Photon.*, vol. 9, pp. 751–757, 2015.

[18] L. E. Nelson, R. M. Jopson, H. Kogelnik, and J. P. Gordon, "Measurement of polarization mode dispersion vectors using the polarization-dependent signal delay method," *Opt. Express*, vol. 6, pp. 158–167, 2000.

[19] G. Milione, D. A. Nolan, and R. R. Alfano, "Determining principal modes in a multimode optical fiber using the mode dependent signal delay method," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 32, pp. 143–149, 2015.

[20] J. Yang and D. A. Nolan, "Using state tomography for characterizing input principal modes in optically scattering medium," *Opt. Express*, vol. 24, pp. 27691–27701, 2016.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[22] C. A. Fuchs, M. C. Hoang, and B. C. Stacey, "The SIC question: History and state of play," *Axioms*, vol. 6, 2017, Art. no. 21.

[23] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan, "A new proof for the existence of mutually unbiased bases," *Algorithmica*, vol. 34, pp. 512–528, 2002.

[24] G. P. Agrawal, *Fiber-Optic Communication Systems*, 2nd ed. New York, NY, USA: Wiley, 1992.

[25] D. Aerts and M. S. de Bianchi, "The extended Bloch representation of quantum mechanics and the hidden-measurement solution to the measurement problem," *Ann. Phys.*, vol. 351, pp. 975–1025, 2014.

[26] J. G. Proakis and M. Salehi, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2008.

[27] G. Strang, *Linear Algebra and Its Applications*. Pacific Grove, CA, USA: Brooks/Cole, 2006.

[28] B. Huttner, C. Geiser, and N. Gisin, "Polarization-induced distortions in optical fiber networks with polarization-mode dispersion and polarization-dependent losses," *IEEE J. Sel. Topics Quantum Electron.*, vol. 6, no. 2, pp. 317–329, Mar./Apr. 2000.

[29] A. Andrusier, M. Shtaif, C. Antonelli, and A. Mecozzi, "Assessing the effects of mode-dependent loss in space-division multiplexed systems," *J. Lightw. Technol.*, vol. 32, no. 7, pp. 1317–1322, Apr. 2014.

[30] K. K. Choutagunta, S. Arik, K. P. Ho, and J. M. Kahn, "Characterizing mode-dependent loss and gain in multimode components," *J. Lightw. Technol.*, vol. 36, no. 18, pp. 3815–3823, Sep. 2018.

[31] J. von Hoyningen-Huene, R. Ryf, and P. Winzer, "LCoS-based mode shaper for few-mode fiber," *Opt. Express*, vol. 21, no. 15, pp. 18097–18110, Jul. 2013.

[32] E. Süli and D. F. Mayers, *An Introduction to Numerical Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[33] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. New York, NY, USA: Cambridge Univ. Press, 2007.

[34] *Mathematica, Version 11.3*, Wolfram Research, Inc., Champaign, IL, USA, 2018.

[35] I. Roudas and J. Kwapisz, "Accurate modal dispersion measurements using maximally-orthogonal Stokes vectors," in *Proc. Conf. Lasers Electro-Opt.*, San Jose, CA, USA, 2018, Paper SM3C.3.

[36] I. Roudas, "Research database," 2018. [Online]. Available: www.montana.edu/opticomm/ResearchDatabase.html

Ioannis Roudas received the B.S. degree in physics and the M.S. degree in electronics and radio engineering from the University of Athens, Athens, Greece, in 1988 and 1990, respectively, and the M.S. and Ph.D. degrees in coherent optical communication systems from the Ecole Nationale Supérieure des Télécommunications (currently Télécom ParisTech), Paris, France, in 1991 and 1995, respectively.

During 1995–1998, he was with the Optical Networking Research Department, Bell Communications Research (Bellcore), Red Bank, NJ, USA. At the same time, he taught for two semesters, as an Adjunct Professor, with Columbia University. He was subsequently with the Photonic Modeling and Process Engineering Department, Corning Inc., from 1999 to 2002. During 2003–2011, he was an Associate Professor of optical communications with the Department of Electrical and Computer Engineering, University of Patras. In addition, he taught, as an Adjunct Professor, with the City University of New York and the Hellenic Open University. During 2011–2016, he was a Research Associate with the Science and Technology Division, Corning, Inc., Corning, NY, USA. Since July 2016, he has been the Gilhousen Telecommunications Chair Professor with the Department of Electrical and Computer Engineering, Montana State University, Bozeman, MT, USA. He has authored or coauthored more than 100 papers in scientific journals and international conferences and holds five patents.

Dr. Roudas is an Associate Editor for the IEEE PHOTONICS JOURNAL.

Jaroslaw (Jarek) Kwapisz received the M.S. degree from the University of Warsaw, Warsaw, Poland, in 1991, and the Ph.D. degree from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 1995. He is a Polish–American mathematician with background in theoretical dynamical systems. He has worked on problems in several subject areas, including integral and differential equations, iterated maps modeling coupled nonlinear oscillators, pattern formation in fourth-order Hamiltonian systems, ergodic theory and entropy in smooth and symbolic dynamics, cohomological Conley index and cocyclic subshifts, almost-periodic tiling spaces and quasi-crystals, Abelian-Nielsen classes and geometry of translation surfaces, and conformal dimension of fractal sets. His current research interests include quasi-symmetric renormalization for infinitely ramified fractals, Anosov maps on infra-nil manifolds, non-Meyer substitution Delone sets, and applications of geometry of Stokes space to multimode fiber-optic communication.

Daniel A. Nolan received the Ph.D. degree in physics from Pennsylvania State University, State College, PA, USA, in 1974. He is currently a Corporate Research Fellow with Corning Research and Development Corporation, Corning, NY, USA. His research activities have included the propagation of light in single-mode and multimode fibers, polarization optics, quantum optics, nonlinear effects in fibers, fiber-optic sensors, and components for local area networks. He holds 101 U.S. patents. He is a Fellow of the Optical Society of America. He is a recipient of a number of awards including the IR100 Award for the invention of infrared polarizing glass, Polarcor, in 1986, the Corning's Outstanding Publication Award in 1989, the Corning's Stookey Award for Exploratory Research in 1995, the Penn State Outstanding Alumni Award from the College of Science for contributions to optical communications in 2001, and the JOURNAL OF LIGHTWAVE TECHNOLOGY Recognition Award in 2008.