Delay Spread Reduction in Mode-Division Multiplexing: Mode Coupling Versus Delay Compensation

Sercan Ö. Arık, Keang-Po Ho, *Senior Member, IEEE*, and Joseph M. Kahn, *Fellow, IEEE*

***Abstract*—Reduction of the group delay (GD) spread is crucial for minimizing signal processing complexity in mode-division multiplexing. Strong mode coupling and GD compensation (concatenating different fibers with opposing GD ordering) are two approaches for reducing the end-to-end GD spread. In this paper, we study the GD behavior in systems where mode coupling and GD compensation are both present. Using a propagation model in generalized Stokes space, we describe the evolution of the GD variance by coupled differential equations. By integration of these equations, we evaluate the GD variance in GD compensated systems with different mode coupling lengths and GD compensation lengths. When the mode coupling length is much longer than the GD compensation length, a low GD variance can be obtained as a result of GD compensation. By contrast, when the mode coupling length is much shorter than the GD compensation length, GD compensation becomes ineffective, but a low GD variance can be obtained as a result of strong mode coupling. The largest GD variance is obtained when the mode coupling length is comparable to the GD compensation length.**

***Index Terms*—Group delay compensation, MIMO signal processing, mode coupling, mode-division multiplexing, multi-mode fibers, space-division multiplexing.**

I. Introduction

As single-mode fiber (SMF) systems approach fundamental capacity limits [1], continued traffic growth motivates mode-division multiplexing (MDM) in multi-mode fiber (MMF) [2]. In MDM, the total capacity per fiber is ideally proportional to the total number of propagating modes D [3], [4]. MDM enables integration of fibers, amplifiers, switches and other components, which facilitates economical and energy-efficient optical network scaling [5], [6]. The integration in MDM comes at the expense of potential challenges, however. These include: (i) crosstalk caused by mode coupling, (ii) distortion caused by modal dispersion and (iii) capacity loss and outage caused by mode-dependent loss/gain (MDL) [4]. An MDM receiver must employ multi-input multi-output (MIMO) signal processing to compensate for (i) and (ii) and achieve a low error probability, while tracking fast changes in the channel [7]–[9]. The cost and power consumption of MIMO signal processing are of concern because of the $D \times D$ matrix structure and a potentially large group delay (GD) spread. Managing the end-to-end GD spread of an MDM system is crucial for making MDM a practical reality.

In some MDM systems, a sufficiently low GD can be obtained by optimization of the fiber index profile. Early MDM experiments used step-index fibers supporting two mode groups ($D = 6$), where a low GD spread can be realized by choosing a core radius at which the GD-versus-radius curves for the two mode groups intersect [10]. This approach does not scale easily beyond two mode groups, since the curves for different pairs of mode groups intersect at different radii. Graded-index fibers with large cores ($D \to \infty$) are known to have very low GD spreads [11]. For $D \sim 12$–30, although the GD spread can be minimized using a graded-index core with depressed cladding [7], [12], [13], it is not sufficiently low to realize practical MIMO signal processing [7]–[9].

Apart from optimization of the fiber index profile, two approaches for minimizing the GD spread are (i) increasing the mode coupling strength to introduce *strong mode coupling* [14] and (ii) concatenating segments of different fiber types that have opposing GD ordering, a method known as *GD compensation* [12], [13], [15]–[18].

In the strong-coupling regime, GD accumulation is minimized by continuous intermixing as the modes propagate, and the GD spread scales with the square-root of propagation distance [14], [19]. As the coupling length is reduced, the end-to-end GD spread asymptotically approaches zero. Since the end-to-end GD spread statistics are governed only by the root-mean-square (r.m.s.) uncoupled GD spread of each fiber section [14], the impact of non-idealities in fiber manufacturing can be reduced. Strong mode coupling is already used to minimize polarization mode dispersion (PMD) in SMF [20].

In typical MMFs used for MDM, random perturbations cause strong coupling within mode groups (*intragroup coupling*) over distances less than 1 km [2], [21]. The intragroup coupling strength can be increased further by intentionally introducing distributed perturbations, e.g., by spinning the MMF during the pulling process [22], similar to the spinning used to reduce PMD in SMF [20]. On the other hand, coupling between modes in different groups (*intergroup coupling*) may not occur even after 100 km of propagation [2], [21], owing to insufficient perturbation of the refractive index. It is desired to obtain substantial intergroup coupling with low loss. An ideal approach would be

Manuscript received April 1, 2015; revised June 23, 2015 and August 26, 2015; accepted August 28, 2015. Date of publication August 31, 2015; date of current version September 25, 2015. This work was supported by a Google Research Award and by Corning, Inc.

S. Ö. Arık and J. M. Kahn are with E. L. Ginzton Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: soarik@stanford.edu; jmk@ee.stanford.edu).

K.-P. Ho is with SiBEAM and Lattice Semiconductor, Sunnyvale, CA 94085 USA (e-mail: kpho@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2015.2475422

to incorporate an appropriate distributed perturbation to induce intergroup coupling in the transmission fiber. An alternate approach is to insert lumped mode scramblers at regular intervals along the link, e.g., long-period fiber grating (LPFG) scramblers, which use periodic index perturbations to phase match modes having different propagation constants [23]. LPFG scramblers may increase a system's mode-averaged and MDL. Hence, it may be desirable to minimize the number of scramblers used in a system and to seek further GD reduction through other means, such as GD compensation.

In GD compensation, GD accumulation is minimized by cascading segments of MMF in which low- and high-order modes exhibit an opposing ordering of GDs. Ideally, in the absence of mode coupling, the end-to-end GD spread of a system can be made zero. This can be achieved, for example, by concatenating fiber type F_1 of length $L_{C,1}$ and zero-mean GD values $(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,D})$ with fiber type F_2 of length $L_{C,2}$ and zero-mean GD values $(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,D}) = -(L_{C,1}/L_{C,2})(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,D})$.

For realization of MMFs with opposing GD ordering, systematic index profile modifications, such as changes of the core radius and index exponent, have been proposed [13], [18]. In order to minimize the impact of mode coupling, segments of each fiber type should be much shorter than the length for mode coupling caused by random perturbations and splices. Depending on the mode coupling strength, this condition may be difficult to satisfy in practice. Hence, understanding the GD behavior in GD-compensated systems with various strengths of mode coupling is of significant interest, and is the focus of this paper.

The remainder of this paper is as follows. Section II describes a model for MDM systems including GD compensation and mode coupling. Section III presents GD statistics computed based on the model in Section II. Sections IV and V provide discussion and conclusions, respectively. Appendices A and B provide mathematical details for the computations presented in Section III. Appendix C describes optimization of GD compensation using an arbitrary number of fiber types.

II. Modeling GD Compensation and Mode Coupling

In the absence of MDL, and neglecting mode-averaged effects, linear propagation in a fiber can be described by a GD operator [4], [21]. The eigenmodes of the GD operator are the principal modes (PMs), and its eigenvalues are the GDs of the PMs [21], [24].

In the case of SMF ($D = 2$), representations of the GD operator in Jones [25], [26] and Stokes [27] spaces are well known. The PMs correspond to the two principal states of polarization (PSPs), which can be represented by one PSP, since the other PSP is orthogonal to it [25]. The GDs are equal and opposite, so the differential GD (DGD) is sufficient to represent them. Hence, in Jones space, the 2×2 GD operator is fully represented by a single 2×1 Jones vector, whose magnitude and direction are the DGD and one of the PSPs, respectively. This Jones vector, in turn, determines a 3×1 Stokes vector, whose magnitude and direction define the DGD and one of the PSPs, and thus fully represent the GD operator [28].

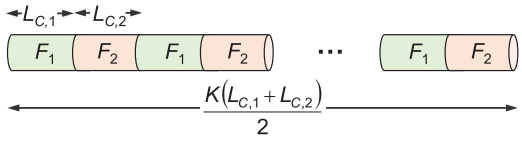

Fig. 1. Multi-segment model for GD-compensated system with fiber types F_1 and F_2, of lengths $L_{C,1}$ and $L_{C,2}$. In total there are K segments, so the total length is $(L_{C,1} + L_{C,2})K/2$.

In the analysis of MMF ($D > 2$), both the generalized Jones [14] and generalized Stokes [29] representations of the GD operator have been used. The generalized Stokes representation facilitates the modeling of mode coupling even when the coupling length is greater than the system length, and enables study of the evolution of the GD continuously along the propagation path. In generalized Jones space, the GD operator can be represented by a $D \times D$ zero-trace Hermitian matrix [14], [21]. Equivalently it may be represented by generalized Jones vectors whose direction and magnitude are, respectively, the PMs and their GDs. In generalized Stokes space, the GD operator can be represented as a $\left(D^2 - 1\right) \times 1$ generalized Stokes vector [29], as explained in Appendix A.

Evolution of the GD operator in generalized Stokes space is described by a stochastic differential equation [29]

$$\frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\partial \boldsymbol{\beta}}{\partial \omega} + \boldsymbol{\beta} \times \boldsymbol{\tau} \tag{1}$$

where $\boldsymbol{\beta}$ is a $\left(D^2 - 1\right) \times 1$ birefringence vector, $\boldsymbol{\tau}$ is a $\left(D^2 - 1\right) \times 1$ dispersion vector, z is propagation distance and ω is angular frequency. Generalizing the linear birefringence model from SMF [27], the z dependence of the birefringence vector can be expressed as

$$\boldsymbol{\beta}(z) = \frac{\partial \boldsymbol{\beta}(z)}{\partial \omega} \cdot \omega + \mathbf{g}(z). \tag{2}$$

(The dependence of $\boldsymbol{\beta}$ and $\partial\boldsymbol{\beta}/d\omega$ on z are suppressed in the remainder of the paper for notational simplicity.) The $\partial\boldsymbol{\beta}/d\omega$ term in (2) represents the deterministic GD determined by the types of fiber employed. The second term $\mathbf{g}(z)$ in (2) is a $\left(D^2 - 1\right)$-dimensional additive noise vector representing random mode coupling. Frequency-independent deterministic terms in (2) are ignored for simplicity, since they do not affect the GD statistics under the mode coupling model (2), in which the perturbations added to birefringence vector are spatially uncorrelated along propagation axis [29], [30].

A GD-compensated system comprises fiber segments with opposing GD ordering such that the corresponding $\partial\boldsymbol{\beta}/d\omega$ terms in (2) compensate each other. Here, we consider GD compensation using two fiber types F_1 and F_2 with lengths $L_{C,1}$ and $L_{C,2}$, as shown in Fig. 1, with deterministic GDs given by

$$\frac{\partial \boldsymbol{\beta}}{\partial \omega} = \begin{cases} \mathbf{b}_1 & z - \left\lfloor \dfrac{z}{L_{C,1} + L_{C,2}} \right\rfloor \quad \in [0, L_{C,1}) \\ \mathbf{b}_2 & z - \left\lfloor \dfrac{z}{L_{C,1} + L_{C,2}} \right\rfloor \quad \in [L_{C,1}, L_{C,1} + L_{C,2}). \end{cases} \tag{3}$$

In (3), $\mathbf{b}_1$ and $\mathbf{b}_2$ are the generalized Stokes vectors corresponding to uncoupled propagation in fiber types F_1 and F_2, such that $\mathrm{diag}\left(\tau_{1,1},\ldots,\tau_{1,D}\right)=\left(b_{1,0}\mathbf{I}+\sum_{i=1}^{D^2-1}b_{1,i}\mathbf{\Lambda}_i\right)/D$ and $\mathrm{diag}\left(\tau_{2,1},\ldots,\tau_{2,D}\right)=\left(b_{2,0}\mathbf{I}+\sum_{i=1}^{D^2-1}b_{2,i}\mathbf{\Lambda}_i\right)/D$, where $\left(\tau_{1,1},\tau_{1,2},\ldots,\tau_{1,D}\right)$ and $\left(\tau_{2,1},\tau_{2,2},\ldots,\tau_{2,D}\right)$ are the uncoupled GDs per unit length. Using (16) and (18), we have $\sqrt{\|\mathbf{b}_1\|^2}/D=\sqrt{\sum_{i=1}^{D}\tau_{1,i}^2/D}$ and $\sqrt{\|\mathbf{b}_2\|^2}/D=\sqrt{\sum_{i=1}^{D}\tau_{2,i}^2/D}$, i.e., the modulus of $\partial\boldsymbol{\beta}/d\omega$ gives the r.m.s. uncoupled GD per unit length scaled by D.

In SMF ($D=2$), the three-dimensional random perturbation $\mathbf{g}\left(z\right)$ is assumed to be distributed as $N\left(\mathbf{0},\left(1/h\right)\mathbf{I}\right)$, where $N\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ is a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and $\mathbf{W}\left(l\right)=\int_0^l\sqrt{h}\mathbf{g}\left(z\right)dz$ is a three-dimensional standard Wiener process (Brownian motion) [27], such that $\mathbf{W}\left(0\right)=\mathbf{0}$, and $\mathbf{W}\left(z\right)\sim N\left(\mathbf{0},z\mathbf{I}\right)$. The *coupling length* h, which is the inverse of the variance of the random perturbation $\mathbf{g}(z)$, describes the strength of mode coupling. Strong coupling corresponds to $z>>h$, while weak coupling corresponds to $z<<h$ [27].

The coupling between modes induced by random perturbations depends on phase matching, so it depends on the propagation constant mismatch between the modes and the longitudinal spatial frequency spectrum of the perturbations [4], [30]. In MMF, most random perturbations are more efficient in coupling degenerate or nearly degenerate modes than nondegenerate modes [21]. To take this into account while generalizing the SMF coupling model in the simplest possible way, we define different *intragroup* and *intergroup coupling lengths*, h_{intra} and h_{inter}, respectively. A similar approach is used to distinguish the strengths of polarization coupling and inter-core spatial coupling in multi-core fibers [31]. Typically, we consider $h_{\mathrm{inter}}>h_{\mathrm{intra}}$. Section III also considers the special case $h_{\mathrm{inter}}=h_{\mathrm{intra}}$, which models mode coupling with just a single parameter. For a D-mode fiber, we assume the $\left(D^2-1\right)$-dimensional $\mathbf{g}(z)$ is distributed as $\mathbf{g}\left(z\right)\sim N\left(\mathbf{0},\mathbf{H}\right)$ where $\mathbf{H}$ is a positive-definite real-valued covariance matrix, and $\mathbf{W}\left(l\right)=\int_0^l\mathbf{H}^{-0.5}\mathbf{g}\left(z\right)dz$ is a $\left(D^2-1\right)$-dimensional Wiener process, such that $\mathbf{W}\left(0\right)=\mathbf{0}$, and $\mathbf{W}\left(z\right)\sim N\left(\mathbf{0},z\mathbf{I}\right)$. To further simplify the mode coupling model, we will assume that $\mathbf{H}$ is diagonal with diagonal elements of either $1/h_{\mathrm{intra}}$ or $1/h_{\mathrm{inter}}$, whose ordering depends on the choice of the basis matrices. Using the generalized Pauli matrix basis described in Appendix A, for $D=6$ (where there are two groups of two and four modes):

$$H_{k,k}=\begin{cases}1/h_{\mathrm{intra}}, & k\in[1,9]\cup[18,35]\\ 1/h_{\mathrm{inter}}, & k\in[10,17]\end{cases}\tag{4}$$

and for $D=12$ (where there are three groups of two, four and six modes):

$$H_{k,k}=\begin{cases}1/h_{\mathrm{intra}}, & k\in[1,18]\cup[27,30]\cup[55,58]\cup[71,143]\\ 1/h_{\mathrm{inter}}, & k\in[19,26]\cup[31,54]\cup[59,70]\,.\end{cases}\tag{5}$$

The evolution of the GD operator (1) and (2) can be represented compactly as a stochastic differential equation

$$\partial\boldsymbol{\tau}=\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}+\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\omega\times\boldsymbol{\tau}\right)\partial z-\boldsymbol{\tau}\times\mathbf{H}^{0.5}\partial\mathbf{W}.\tag{6}$$

Note that the (6) is valid regardless of whether $\mathbf{H}$ is diagonal or non-diagonal.

III. GD Statistics

GD statistics are governed by the stochastic evolution of the dispersion vector $\boldsymbol{\tau}$ (6). The Hermitian matrix corresponding to $\boldsymbol{\tau}$, $\left(\tau_0\mathbf{I}+\sum_{i=1}^{D^2-1}\tau_i\mathbf{\Lambda}_i\right)/D$ is the GD operator, whose eigenvectors are the PMs and whose eigenvalues are the coupled GDs. In the absence of random perturbations, the generalized Stokes vector $\boldsymbol{\tau}$ is deterministic and is equal to $z\partial\boldsymbol{\beta}/\partial\omega$, which yields $\boldsymbol{\tau}/z=\boldsymbol{\beta}/\omega=\partial\boldsymbol{\beta}/\partial\omega$, i.e., the three vectors $\boldsymbol{\tau}$, $\boldsymbol{\beta}$ and $\partial\boldsymbol{\beta}/\partial\omega$ are aligned. In this case, the GDs are equal to the uncoupled GDs, which accumulate linearly with fiber length. In the presence of random perturbations, the vectors $\boldsymbol{\tau}$ and $\partial\boldsymbol{\beta}/\partial\omega$ become misaligned from each other, depending on the coupling strength. The direction and modulus of $\boldsymbol{\tau}$ evolves stochastically along the length of the fiber. The coupled GDs become random variables.

As explained in Section IV, in a system comprising many fiber spans (typically of length $L_{amp}=80-100\,\mathrm{km}$ in terrestrial systems), strong coupling can be induced by including one mode scrambler per span. The end-to-end peak-to-peak GD spread that governs MIMO signal processing complexity is determined by the r.m.s. GD per span and the total number of spans [8], [9]. Hence, the r.m.s. GD after a span of one or more fibers is of interest. The r.m.s. GD can be conveniently represented as $\sqrt{\boldsymbol{\tau}\cdot\boldsymbol{\tau}}/D=\sqrt{\|\boldsymbol{\tau}\|^2/D}$ (see [29] or Appendix A), which is independent of the direction of $\boldsymbol{\tau}$ and depends on $\boldsymbol{\tau}$ only through its modulus.

Evolution of the probability density function of $\boldsymbol{\tau}$ along z is described by Fokker–Planck equations [26], [32], and the expected value of a function of $\boldsymbol{\tau}$ can be described by a set of deterministic ordinary differential equations. The deterministic differential equation for $E\left\{\|\boldsymbol{\tau}\|^2\right\}$ (where $E\{\}$ denotes expectation) is derived in [31] and is re-derived in Appendix B using a different approach (see (41)):

$$\frac{\partial}{\partial z}E\left\{\|\boldsymbol{\tau}\|^2\right\}=E\left\{2\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right\}=2\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot E\left\{\boldsymbol{\tau}\right\}.\tag{7}$$

The deterministic differential equation for $E\left\{\boldsymbol{\tau}\right\}$ is derived in Appendix B (see (50) and (51)):

$$\frac{\partial}{\partial z}E\left\{\boldsymbol{\tau}\right\}=\frac{\partial\boldsymbol{\beta}}{\partial\omega}-\mathbf{Q}\cdot E\left\{\boldsymbol{\tau}\right\}\tag{8}$$

where the elements of the $\left(D^2-1\right)\times\left(D^2-1\right)$ matrix $\mathbf{Q}$ are given by

$$\mathbf{Q}_{j,l}=\frac{1}{2}\sum_{m,k=1}^{D^2-1}f_{k,m,l}f_{k,m,j}H_{k,k}.\tag{9}$$

Modification of $\mathbf{Q}$ for the general case beyond the assumed mode coupling model is also described in Appendix B. In the

special case $\mathbf{H} = (1/h)\,\mathbf{I}$ ($h_{\text{inter}} = h_{\text{intra}} = h$), $\mathbf{Q} = (1/h)\,\mathbf{I}$, and using property (24), expression (8) simplifies to

$$\frac{\partial}{\partial z} E\{\boldsymbol{\tau}\} = \frac{\partial \boldsymbol{\beta}}{\partial \omega} - \frac{1}{h} E\{\boldsymbol{\tau}\}. \tag{10}$$

Analytical solutions of the linear equations (7), (8) and (10) can be obtained when $\partial\boldsymbol{\beta}/\partial\omega$ is a simple function of z.

We focus first on the case where $\partial\boldsymbol{\beta}/\partial\omega$ is constant, which corresponds to a system comprising a single fiber type. Assuming an initial condition $E\{\boldsymbol{\tau}\} = \mathbf{0}$ at $z = 0$, integration of (8) yields

$$E\{\boldsymbol{\tau}\} = \mathbf{Q}^{-1}\left(\mathbf{I} - \exp(-z\mathbf{Q})\right)\frac{\partial \boldsymbol{\beta}}{\partial \omega}. \tag{11}$$

Using (11) and assuming an initial condition $E\left\{\|\boldsymbol{\tau}\|^2\right\} = 0$ at $z = 0$, integration of (7) yields

$$E\left\{\|\boldsymbol{\tau}\|^2\right\} = 2\frac{\partial \boldsymbol{\beta}}{\partial \omega} \cdot \left(\mathbf{Q}^{-1} z - \mathbf{Q}^{-2}\left(\mathbf{I} - \exp(-z\mathbf{Q})\right)\right)\frac{\partial \boldsymbol{\beta}}{\partial \omega}. \tag{12}$$

Expression (12) is also given in [31], where the matrix $\mathbf{Q}$ is defined in terms of a cross correlation of the perturbation vector. In the special case $\mathbf{H} = (1/h)\,\mathbf{I}$ with $h_{\text{inter}} = h_{\text{intra}} = h$, we have $\mathbf{Q} = (1/h)\,\mathbf{I}$ and (11) simplifies to

$$E\{\boldsymbol{\tau}\} = h\left(1 - e^{-z/h}\right)\frac{\partial \boldsymbol{\beta}}{\partial \omega} \tag{13}$$

and the squared modulus (12) simplifies to [27]

$$E\left\{\|\boldsymbol{\tau}\|^2\right\} = 2\left\|\frac{\partial \boldsymbol{\beta}}{\partial \omega}\right\|^2 \left(hz - h^2\left(1 - e^{-z/h}\right)\right). \tag{14}$$

In the weak-coupling regime, where $h >> z$ and $h^2 e^{-z/h} \approx h^2 - zh + z^2/2$, the r.m.s. GD is $\sqrt{E\left\{\|\boldsymbol{\tau}\|^2\right\}/D^2} \approx \sqrt{\|\partial\boldsymbol{\beta}/\partial\omega\|^2/D^2}\,z$, which is proportional to the r.m.s. uncoupled GD per unit length, and scales linearly with propagation length [27]. In the strong-coupling regime, where $h << z$ and $hz + h^2 e^{-z/h} - h^2 \approx hz$, the r.m.s. GD is $\sqrt{E\left\{\|\boldsymbol{\tau}\|^2\right\}/D^2} \approx \sqrt{2\|\partial\boldsymbol{\beta}/\partial\omega\|^2 hz/D^2}$, i.e., it is proportional to r.m.s. uncoupled GD per unit length and scales with $\sqrt{z}$ [27]. As a side note, we can relate the generalized Stokes and generalized Jones representations using the ratio between the r.m.s. GDs in the strong- and weak-coupling regimes, which is $\sqrt{2h/z}$. In the generalized Jones representation, the coupling strength is described by a section length L_{sec} in a multi-section model [14], which corresponds to the length over which a propagating electric field remains correlated. The ratio between the r.m.s. GDs in the strong- and weak-coupling regimes is $\sqrt{L_{\text{sec}}/z}$ [14]. Equating the r.m.s. GD reduction ratios found in the two models, we find that the section length in the multi-section generalized Jones model [14] is twice the coupling length in the generalized Stokes model, $L_{\text{sec}} = 2h$.

We now focus on the case where $\partial\boldsymbol{\beta}/\partial\omega$ is a piecewise constant function of z, which describes a system with GD compensation. We assume the model in Fig. 1 such that each span comprises K segments of fiber types F_1 and F_2 and $\partial\boldsymbol{\beta}/\partial\omega$ takes the form (3). The method of solution for $E\{\boldsymbol{\tau}\}$ in this case differs in three ways from (11) because $\partial\boldsymbol{\beta}/\partial\omega$ is piecewise constant and $E\{\boldsymbol{\tau}\}$ is continuous. First, different values of $\partial\boldsymbol{\beta}/\partial\omega$ for different fiber segments are used in (11). Second, in each fiber segment, the function (11) is computed at $(z - z^*)$ instead of z, where z^* is the end point of the previous fiber segment. Third, a term $\exp\left(-(z - z^*)\,\mathbf{Q}\right)(E\{\boldsymbol{\tau}\})^*$ is added to (11) to account for a nonzero initial condition, where $(E\{\boldsymbol{\tau}\})^*$ denotes the value of $E\{\boldsymbol{\tau}\}$ at the end of the previous fiber segment. For example, for $z \in [L_{C,1}, L_{C,1} + L_{C,2})$, the expression for $E\{\boldsymbol{\tau}\}$ becomes $\mathbf{Q}^{-1}\left(\mathbf{b}_2 + \exp\left(-(z - L_{C,1})\,\mathbf{Q}\right)(\mathbf{b}_1 - \mathbf{b}_2) - \exp(-z\mathbf{Q})\mathbf{b}_1\right)$. Since analytical expressions become unwieldy as the number of fiber segments increases, numerical integration of (7), (8) and (10) is preferable in most cases of practical interest.

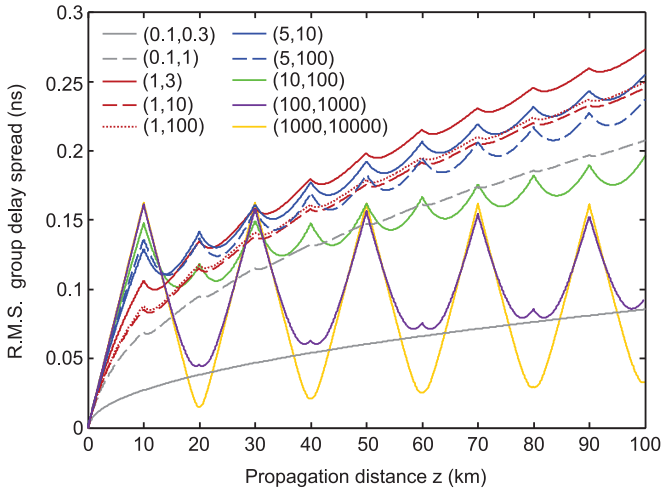

Fig. 2. R.M.S. GD spread versus propagation distance z for $D = 12$ modes in a 100-km fiber span comprising ten GD compensation segments of length $L_{C,1} = L_{C,2} = 10$ km. The legends correspond to different intra- and intergroup coupling lengths $(h_{\text{intra}}, h_{\text{inter}})$ in kilometer.

In computing numerical results, we consider $D = 12$ modes and GD values in $\mathbf{b}_1$ corresponding to a graded-index graded depressed-cladding (GIGDC) fiber [7], which has uncoupled GD values of (17.6, 17.6, 7.6, 7.6, 15.6, 15.6, −27.9, −27.9, −14.4, −14.4, 1.5, 1.5) ps/km. We assume $\mathbf{b}_1 = -\mathbf{b}_2$ and $L_{C,1} = L_{C,2}$ for precise GD compensation. Numerical analysis of the GIGDC fiber and a compensating fiber designed following [13], [18] confirms that the mode group degeneracies and the ordering of the propagation constants are preserved while the ordering of the GDs is reversed. We assume a span length of $K(L_{C,1} + L_{C,2})/2 = 100$ km. Figs. 2 and 3 show the evolution of the r.m.s. GD with propagation distance z, assuming compensation segment lengths $L_{C,1} = L_{C,2}$ of 10 and 5 km, respectively, for different values of intragroup and intergroup coupling lengths $(h_{\text{intra}}, h_{\text{inter}})$.

We also consider the special case $h_{\text{inter}} = h_{\text{intra}}$ using the differential equations (7) and (11). Fig. 4 shows the evolution of the r.m.s. GD along z, assuming compensation segment lengths $L_{C,1} = L_{C,2} = 10$ km for different values of $h_{\text{inter}} = h_{\text{intra}}$.

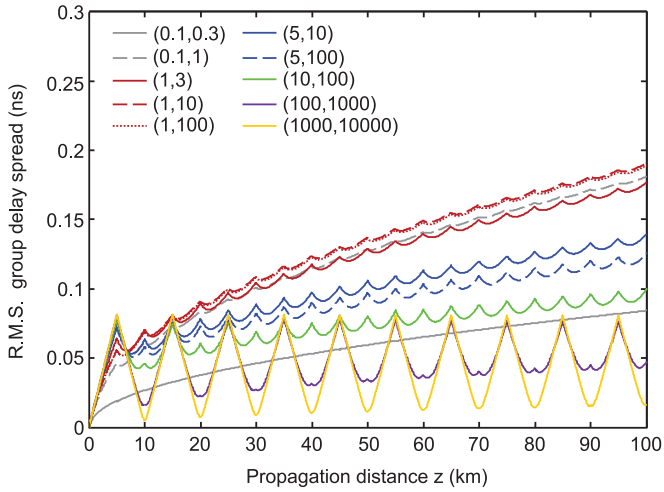

Fig. 3. R.M.S. GD spread versus propagation distance z for $D = 12$ modes in a 100-km fiber span comprising 20 GD compensation segments of length $L_{C,1} = L_{C,2} = 5$ km. The legends correspond to different intra- and intergroup coupling lengths $(h_{\mathrm{intra}}, h_{\mathrm{inter}})$ in kilometer.

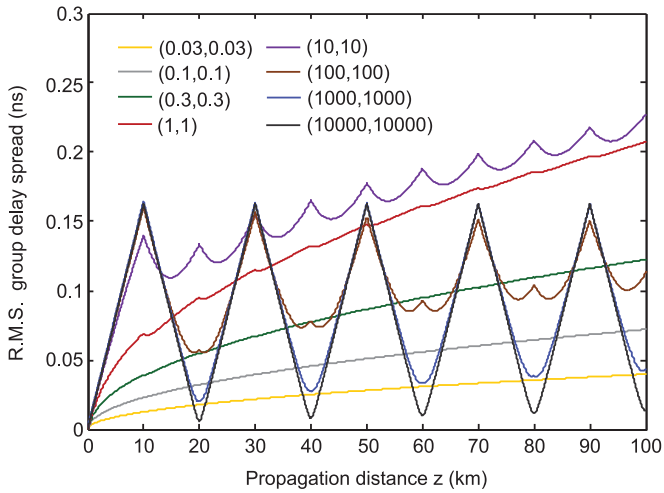

Fig. 4. R.M.S. GD spread versus propagation distance z for $D = 12$ modes in a 100-km fiber span comprising ten GD compensation segments of length $L_{C,1} = L_{C,2} = 10$ km. The legends correspond to different intra- and intergroup coupling lengths $(h_{\mathrm{intra}}, h_{\mathrm{inter}})$ in kilometer, where $h_{intra} = h_{\mathrm{inter}}$.

In the weak-coupling regime, i.e., $h_{\mathrm{inter}}, h_{\mathrm{intra}} >> L_{C,1} = L_{C,2}$, the r.m.s. GD accumulates almost linearly with z. Since opposing GDs in consecutive segments cancel each other, a very small overall r.m.s. GD is obtained. For the values of h_{inter} and h_{intra} considered, the overall r.m.s. GD is reduced by decreasing $L_{C,1} = L_{C,2}$.

In the strong-coupling regime, i.e., $h_{\mathrm{inter}}, h_{\mathrm{intra}} << L_{C,1} = L_{C,2}$, the observed r.m.s. GD is consistent with results obtained from the multi-section generalized Jones matrix model [14]. The r.m.s. GD accumulates roughly in proportion to $\sqrt{z}$, and is roughly proportional to $\sqrt{h_{\mathrm{inter}}}$. GD cancellation is not observed, because in each compensation segment of length $L_{C,1} = L_{C,2}$, the GD accumulation depends only on the r.m.s. uncoupled GDs in fiber types F_1 and F_2, and not on the signs of the uncoupled GDs [14]. Hence, the overall r.m.s. GD does not depend on $L_{C,1} = L_{C,2}$, and is reduced by decreasing h_{inter}.

In the intermediate-coupling regime, i.e., where h_{inter} and h_{intra} are comparable to $L_{C,1} = L_{C,2}$, the r.m.s. GD accumulates sublinearly with z, and partial GD cancellation is observed. For a given intergroup coupling length h_{inter}, decreasing the intragroup coupling length h_{intra} may decrease or increases the r.m.s. GD, depending on whether GD compensation or mode coupling is dominant. The intermediate-coupling regime yields the highest overall r.ms. GD values.

IV. Discussion

The results of Section III show that careful engineering of a transmission link and its components is required to minimize the end-to-end GD spread of the link.

When mode coupling is very weak, GD compensation is highly effective in minimizing the end-to-end r.m.s. GD. For the values considered here, this requires h_{inter} to be three to four orders of magnitude larger than $L_{C,1}$ and $L_{C,2}$, and also requires h_{intra} to be two to three orders of magnitude larger than $L_{C,1}$ and $L_{C,2}$, since modes within a group do not generally have degenerate GDs. To satisfy these conditions, short compensation lengths $L_{C,1}$ and $L_{C,2}$ should be used, and mode coupling within compensating fiber segments should be minimized. Furthermore, two types of fibers with nearly precisely opposing GD profiles, $\tau_{2,i} \approx -(L_{C,1}/L_{C,2})\,\tau_{1,i}, \quad 1 \le i \le D$ are required. In practice, for $D > 6$, obtaining precise GD cancellation using two, or even more, types of fiber is challenging [33]. When precise GD cancellation is not possible, the compensation lengths can be optimized to minimize the GD spread. Appendix C describes a procedure to optimize the compensation lengths for a general number of fiber types.

When mode coupling is strong, GD compensation becomes ineffective, and the ordering of GD values becomes immaterial. For the values considered here, this requires both the coupling lengths h_{inter} and h_{intra} to be three to four orders of magnitude smaller than both $L_{C,1}$ and $L_{C,2}$. In the strong-mode-coupling regime, the r.m.s. GD accumulates in proportion to $\sqrt{z}$, and decreases with an increase in coupling strength. To achieve a low r.m.s. GD using mode coupling, strong perturbations should be introduced within the fiber segments, which may increase mode-averaged and MDL.

In order to control the complexity of cable manufacturing and installation, the compensation lengths $L_{C,1}$ and $L_{C,2}$ presumably cannot be much smaller than the splice spacing in current systems, which is about 1–5 km. Experiments [2] and estimation [21] suggest that typically, h_{intra} is of order 0.1–1 km, while h_{inter} is of order 10–100 km. Given these values, Figs. 2 and 3 suggest that GD compensation is likely to be only partially effective in reducing the end-to-end GD spread in deployed systems.

Thus far, we have only focused on the GD of a single span comprising multiple GD compensation segments. In a long-haul system comprising multiple spans, the end-to-end GD spread depends on the GD spread in each span and also on the coupling between spans. A promising design strategy is to insert mode scramblers (e.g., LPFGs [23]) between spans, as this can

minimize the accumulation of GD spread [14], MDL [34] and nonlinear impairments [35–[37]. In a system using mode scramblers between spans, the end-to-end peak-to-peak GD spread is $\sim 4.5\sqrt{K_{\mathrm{amp}}}$ times the r.m.s. GD spread per span, where K_{amp} is the number of spans, assuming $K_{\mathrm{amp}} >> 1$ [9]. For example, consider a system employing 20 GD-compensated spans, each 100 km long, with mode scrambling between spans. Assuming compensation lengths of 5–10 km and $D = 12$ modes, and using the values in Figs. 2 and 3, the end-to-end peak-to-peak GD spread should not exceed ~6 ns. Assuming a symbol rate of 35 Gbaud, this corresponds to ~210 symbol intervals, which is sufficiently small to render the complexity of adaptive MIMO signal processing acceptable [8].

V. Conclusion

In MDM systems, minimization of the GD spread is crucial to render MIMO signal processing complexity acceptable. We studied the behavior of the GD spread in systems having both GD compensation and mode coupling. Using a propagation model in generalized Stokes space, we derived ordinary differential equations describing the evolution of the GD variance. GD compensation is found to be highly effective only when both intra- and intergroup coupling lengths are far larger than the lengths of GD compensating segments, a condition that may be difficult to achieve in deployed systems. Combining intra-span GD compensation with inter-span mode scrambling may achieve acceptably low GD spread in practice.

Appendices

A. Stokes Space Operator Representations

Here, we review some operator representations in generalized Stokes space.

The GD operator may be represented in generalized Jones space by a $D \times D$ zero-trace Hermitian matrix $\mathbf{M}$. Alternatively, $\mathbf{M}$ may be represented in generalized Jones space by its eigenvectors (the PMs) and its eigenvalues (the GDs of the PMs). In generalized Stokes space, the GD operator may be represented in a basis of matrices $\mathbf{\Lambda}_i$ $(1 \le i \le D^2 - 1)$. A convenient basis is the set of generalized Pauli matrices [29], which are also the generator matrices for the special unitary (SU) group in its Lie algebra up to a simple scaling [38]. The generalized Pauli matrices are traceless:

$$Tr\left\{\mathbf{\Lambda}_i\right\} = 0 \tag{15}$$

and have the trace orthogonality property

$$\frac{1}{D} Tr\left\{\mathbf{\Lambda}_i \mathbf{\Lambda}_j\right\} = \delta_{i,j} \tag{16}$$

where $\delta_{i,j}$ is the Kronecker delta. Methods to generate generalized Pauli basis matrices are given in [29]. A general representation of $\mathbf{M}$ is given by

$$\mathbf{M} = \frac{1}{D}\left(m_0 \mathbf{I} + \sum_{i=1}^{D^2-1} m_i \mathbf{\Lambda}_i\right) \tag{17}$$

where $\mathbf{I}$ is the identity matrix, and the coefficients are $m_0 = Tr\left\{\mathbf{M}\right\} = 0$ and $m_i = Tr\left\{\mathbf{\Lambda}_i \mathbf{M}\right\}$ [29]. The $\left(D^2-1\right)$-dimensional vector $\mathbf{m} = \left[\, m_1 \, \cdots \, m_{D^2-1} \,\right]$ is the *generalized Stokes space vector* corresponding to $\mathbf{M}$.

Two multiplication operations for Stokes space vectors are the dot and cross products. Let $\mathbf{a}$ and $\mathbf{b}$ be two $\left(D^2-1\right)$-dimensional vectors. The dot product between $\mathbf{a}$ and $\mathbf{b}$ is [29]

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{D^2-1} a_i b_i = \frac{1}{D} Tr\left\{\left(\sum_{i=1}^{D^2-1} a_i \mathbf{\Lambda}_i\right)\left(\sum_{i=1}^{D^2-1} b_i \mathbf{\Lambda}_i\right)\right\}. \tag{18}$$

Using (16) and (18), we observe $Tr\left\{\mathbf{M}\mathbf{M}^*\right\} = \left\|\mathbf{m}\right\|^2 / D$. Note that if $\mathbf{M}$ is the GD operator, the GD values are the eigenvalues of $\mathbf{M}$ and the sum of the square of eigenvalues is equal to $Tr\left\{\mathbf{M}\mathbf{M}^*\right\}$ or $\left\|\mathbf{m}\right\|^2 / D$. Hence, the r.m.s. GD is given by $\sqrt{\left\|\mathbf{m}\right\|^2}/D$ [29]. The cross product between $\mathbf{a}$ and $\mathbf{b}$ is [29]

$$\mathbf{a} \times \mathbf{b} = \sum_{i,j,k=1}^{D^2-1} f_{i,j,k} a_i b_j \mathbf{e}_k, \tag{19}$$

where $\mathbf{e}_k$ are the orthogonal unit-length basis vectors of the $\left(D^2-1\right)$-dimensional generalized Stokes space and $f_{i,j,k}$ are structure coefficients given by [29], [39]

$$f_{i,j,k} = \frac{j}{D^2} Tr\left\{\mathbf{\Lambda}_k \left(\mathbf{\Lambda}_i \mathbf{\Lambda}_j - \mathbf{\Lambda}_j \mathbf{\Lambda}_i\right)\right\}. \tag{20}$$

Apart from the factor of $1/D^2$, the structure coefficients are the same as those used in Lie algebra [39]. The structure coefficients have the following properties [39]:

$$f_{k,l,m} = f_{l,m,k} = f_{m,k,l}, \tag{21}$$

$$f_{k,l,m} = -f_{l,k,m} = -f_{k,m,l}, \tag{22}$$

$$f_{k,k,m} = f_{m,k,k} = f_{k,m,k} = 0, \tag{23}$$

$$\sum_{k,l} f_{k,l,m} f_{k,l,n} = 2\delta_{n,m}. \tag{24}$$

Properties (21)–(23) can easily be observed from the definition (20). Property (24) holds because the Killing form for the compact group SU(D) is the Kronecker delta [39]. The cross product $\mathbf{a} \times \mathbf{b}$ can also be represented as $\mathbf{A}\mathbf{b}$, where the elements of the $\left(D^2-1\right) \times \left(D^2-1\right)$ matrix $\mathbf{A}$ are given by $A_{j,k} = \sum_{i=1}^{D^2-1} f_{i,j,k} a_i$. We use this property in Appendix B, while using the cross product operator notation for matrix multiplication.

B. Fokker–Planck Equation Derivations

Consider a $\left(D^2-1\right)$-dimensional stochastic differential equation in the form

$$\partial \mathbf{X} = \boldsymbol{\mu} \partial z + \mathbf{\Sigma} \partial \mathbf{W}, \tag{25}$$

where $\boldsymbol{\mu}$ is a $\left(D^2-1\right)$-dimensional drift vector with elements μ_i corresponding to evolution of the mean of $\mathbf{X}$, $\mathbf{\Sigma}$ is a $\left(D^2-1\right) \times \left(D^2-1\right)$ matrix with elements $\sigma_{i,j}$ and $\mathbf{W}$ is a Wiener process. Evolution of $p\left(\mathbf{x}, z\right)$, the probability density function of $\mathbf{X}$ as a

function of z, is described by a differential equation [32]

$$\frac{\partial}{\partial z}p\left(\mathbf{x},z\right)=\left(\mathbf{G_1}+\mathbf{G_2}\right)p\left(\mathbf{x},z\right). \tag{26}$$

Here, the operator $\mathbf{G_1}$ is

$$\mathbf{G_1}=\nabla_{\mathbf{x}}\cdot\boldsymbol{\mu}=\sum_{k=1}^{D^2-1}\frac{\partial}{\partial x_k}\mu_k, \tag{27}$$

where $\nabla_{\mathbf{x}}$ is the gradient vector with respect to $\mathbf{x}$. Choice of the operator $\mathbf{G_2}$ depends on the interpretation of the stochastic differential equation. There are two widely used stochastic calculi: the Stratonovich and Ito calculi [32]. The Ito interpretation is typically used in biological sciences and mathematical finance, where the future of a stochastic process is considered unknown and a continuous approximation of a discrete system is employed [32]. The Stratonovich interpretation is typically used in physical systems when white noise is used as an idealization of a smooth noise process [27], [32], [40]. Here, we adopt the Stratonovich interpretation, in which the operator $\mathbf{G_2}$ is [27], [32], [40]:

$$\begin{aligned}\mathbf{G}_2 &= \frac{1}{2}\left(\boldsymbol{\Sigma}^T\nabla_{\mathbf{x}}\right)^T\left(\boldsymbol{\Sigma}^T\nabla_{\mathbf{x}}\right)\\ &= \frac{1}{2}\sum_{k=1}^{D^2-1}\sum_{n=1}^{D^2-1}\frac{\partial}{\partial x_n}\sigma_{n,k}\sum_{m=1}^{D^2-1}\frac{\partial}{\partial x_m}\sigma_{m,k}.\end{aligned} \tag{28}$$

Using (23), for a differentiable function f, the evolution of $E\left\{f\left(\mathbf{X}\right)\right\}$ can be expressed as [27], [32]

$$\frac{\partial}{\partial z}E\left\{f\left(\mathbf{X}\right)\right\}=E\left\{\left(\mathbf{G_1}+\mathbf{G_2}\right)f\left(\mathbf{X}\right)\right\}. \tag{29}$$

Considering the stochastic differential equation (6), we have $\mathbf{X}=\boldsymbol{\tau}$, $\boldsymbol{\mu}=\partial\boldsymbol{\beta}/\partial\omega+\omega\partial\boldsymbol{\beta}/\partial\omega\times\boldsymbol{\tau}$ and $\boldsymbol{\Sigma}=-\boldsymbol{\tau}\times\mathbf{H}^{1/2}$. The corresponding $\mathbf{G_1}$ and $\mathbf{G}_2$ operators are

$$\mathbf{G_1}=\nabla_{\boldsymbol{\tau}}\cdot\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}+\omega\frac{\partial\boldsymbol{\beta}}{\partial\omega}\times\boldsymbol{\tau}\right) \tag{30}$$

and

$$\mathbf{G}_2=\frac{1}{2}\left(\left(\boldsymbol{\tau}\times\mathbf{H}^{1/2}\right)^T\nabla_{\boldsymbol{\tau}}\right)^T\left(\left(\boldsymbol{\tau}\times\mathbf{H}^{1/2}\right)^T\nabla_{\boldsymbol{\tau}}\right). \tag{31}$$

Using structure coefficient property (22), we observe that $(\boldsymbol{\tau}\times)^T=-\boldsymbol{\tau}\times$. Furthermore, $\mathbf{H}^T=\mathbf{H}$, since $\mathbf{H}$ is a covariance matrix. Hence, (31) simplifies to

$$\mathbf{G}_2=-\frac{1}{2}\left(\nabla_{\boldsymbol{\tau}}\right)^T\boldsymbol{\tau}\times\mathbf{H}\tau\times\nabla_{\boldsymbol{\tau}}. \tag{32}$$

Using the operators $\mathbf{G_1}$ and $\mathbf{G_2}$, we will derive differential equations for $E\left\{\left\|\boldsymbol{\tau}\right\|^2\right\}$ and $E\left\{\partial\boldsymbol{\beta}/\partial\omega\cdot\boldsymbol{\tau}\right\}$, considering the functions $f(\boldsymbol{\tau})=\left\|\boldsymbol{\tau}\right\|^2$ and $f(\boldsymbol{\tau})=\partial\boldsymbol{\beta}/\partial\omega\cdot\boldsymbol{\tau}$.

We first compute $\mathbf{G_1}$ and $\mathbf{G_2}$ operators for $\left\|\boldsymbol{\tau}\right\|^2$. The $\mathbf{G_1}$ operator yields

$$\mathbf{G_1}\left\|\boldsymbol{\tau}\right\|^2=\nabla_{\boldsymbol{\tau}}\cdot\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}+\omega\frac{\partial\boldsymbol{\beta}}{\partial\omega}\times\boldsymbol{\tau}\right)\left\|\boldsymbol{\tau}\right\|^2. \tag{33}$$

Using the property of structure coefficients (23), it is observed that the variable τ_k does not appear in the kth element of $\partial\boldsymbol{\beta}/\partial\omega\times\boldsymbol{\tau}$, so the order of differentiation can be exchanged:

$$\mathbf{G_1}\left\|\boldsymbol{\tau}\right\|^2=\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}+\omega\frac{\partial\boldsymbol{\beta}}{\partial\omega}\times\boldsymbol{\tau}\right)\cdot\nabla_{\boldsymbol{\tau}}\left\|\boldsymbol{\tau}\right\|^2. \tag{34}$$

Because $\nabla_{\boldsymbol{\tau}}\left\|\boldsymbol{\tau}\right\|^2=2\boldsymbol{\tau}$,

$$\mathbf{G_1}\left\|\boldsymbol{\tau}\right\|^2=2\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}+\omega\frac{\partial\boldsymbol{\beta}}{\partial\omega}\times\boldsymbol{\tau}\right)\cdot\boldsymbol{\tau}. \tag{35}$$

Using structure coefficient properties (22) and (23), it can be shown that the second term of (35) can be expressed as

$$\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\times\boldsymbol{\tau}\right)\cdot\boldsymbol{\tau}=-\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\left(\boldsymbol{\tau}\times\boldsymbol{\tau}\right). \tag{36}$$

Here, the kth component of $\boldsymbol{\tau}\times\boldsymbol{\tau}$ is given by $\sum_{i,j}f_{i,j,k}\tau_i\tau_j$ and is decomposed as

$$\sum_i f_{i,i,k}\tau_i^2+\sum_{\substack{i,j\\i<j}}f_{i,j,k}\tau_i\tau_j+\sum_{\substack{i,j\\i<j}}f_{j,i,k}\tau_i\tau_j. \tag{37}$$

Using the structure coefficient properties (22) and (23), it is observed that (37) yields 0, which can also be interpreted as implying that $\partial\boldsymbol{\beta}/\partial\omega\times\boldsymbol{\tau}$ is perpendicular to $\boldsymbol{\tau}$. Hence, we find that

$$\mathbf{G_1}\left\|\boldsymbol{\tau}\right\|^2=2\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}. \tag{38}$$

The $\mathbf{G}_2$ operator yields

$$\mathbf{G}_2\left\|\boldsymbol{\tau}\right\|^2=-\frac{1}{2}\left(\nabla_{\boldsymbol{\tau}}\right)^T\boldsymbol{\tau}\times\mathbf{H}\tau\times\nabla_{\boldsymbol{\tau}}\left\|\boldsymbol{\tau}\right\|^2. \tag{39}$$

Using $\nabla_{\boldsymbol{\tau}}\left\|\boldsymbol{\tau}\right\|^2=2\boldsymbol{\tau}$, it follows that $\left(\mathbf{H}\tau\times\nabla_{\boldsymbol{\tau}}\right)\left\|\boldsymbol{\tau}\right\|^2=2\mathbf{H}\tau\times\boldsymbol{\tau}$. Since $\boldsymbol{\tau}\times\boldsymbol{\tau}=\mathbf{0}$, as shown above, we have

$$\mathbf{G}_2\left\|\boldsymbol{\tau}\right\|^2=0. \tag{40}$$

Combining operators $\mathbf{G}_1$ and $\mathbf{G}_2$ using (29), we get

$$\frac{\partial}{\partial z}E\left\{\left\|\boldsymbol{\tau}\right\|^2\right\}=E\left\{2\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right\}=2\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot E\left\{\boldsymbol{\tau}\right\}. \tag{41}$$

We now compute $\mathbf{G_1}$ and $\mathbf{G_2}$ operators for $\partial\boldsymbol{\beta}/\partial\omega\cdot\boldsymbol{\tau}$ for a diagonal $\mathbf{H}$. The $\mathbf{G_1}$ operator yields

$$\begin{aligned}\mathbf{G_1}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right) &= \nabla_{\boldsymbol{\tau}}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right)\cdot\frac{\partial\boldsymbol{\beta}}{\partial\omega}\\ &\quad+\omega\nabla_{\boldsymbol{\tau}}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right)\cdot\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\times\boldsymbol{\tau}\right).\end{aligned} \tag{42}$$

Note that $\nabla_{\boldsymbol{\tau}}\cdot\left(\partial\boldsymbol{\beta}/\partial\omega\cdot\boldsymbol{\tau}\right)=\partial\boldsymbol{\beta}/\partial\omega$ and $\partial\boldsymbol{\beta}/\partial\omega\cdot\ \left(\partial\boldsymbol{\beta}/\partial\omega\times\boldsymbol{\tau}\right)=0$ (similar to the argument made for $\boldsymbol{\tau}$ using (36) and (37)). Hence

$$\mathbf{G_1}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right)=\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\frac{\partial\boldsymbol{\beta}}{\partial\omega}=\left\|\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right\|^2. \tag{43}$$

The $\mathbf{G}_2$ operator yields

$$\mathbf{G}_2\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right)=-\frac{1}{2}\left(\nabla_{\boldsymbol{\tau}}\right)^T\boldsymbol{\tau}\times\mathbf{H}\tau\times\nabla_{\boldsymbol{\tau}}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right). \tag{44}$$

Simplification of the first differentiation yields

$$\mathbf{G}_2\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right)=-\frac{1}{2}\left(\nabla_{\boldsymbol{\tau}}\right)^T\boldsymbol{\tau}\times\mathbf{H}\tau\times\frac{\partial\boldsymbol{\beta}}{\partial\omega}. \tag{45}$$

The kth component of $(\mathbf{H}\tau \times \partial\boldsymbol{\beta}/\partial\omega)$ is

$$H_{k,k}\sum_{i,j} f_{i,j,k}\tau_i \left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right)_j. \tag{46}$$

The mth element of the cross product between $\boldsymbol{\tau}$ and $(\mathbf{H}\tau \times \partial\boldsymbol{\beta}/\partial\omega)$ is

$$\sum_{l,k=1}^{D^2-1} f_{l,k,m}\tau_l H_{k,k}\sum_{i,j} f_{i,j,k}\tau_i \left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right)_j. \tag{47}$$

The derivative of (44) with respect to τ_m has non-zero terms only when $l = m$ or $i = m$. However, when $l = m$, $f_{l,k,m}$ becomes 0. Using this simplification, we obtain

$$\mathbf{G}_2\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right) = -\frac{1}{2}\sum_{m=1}^{D^2-1}\sum_{j,k,l} f_{k,m,l}f_{k,m,j}\tau_l H_{k,k}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right)_j. \tag{48}$$

Combining operators $\mathbf{G}_1$ and $\mathbf{G}_2$ using (29) we obtain

$$\frac{d}{dz}E\left\{\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\boldsymbol{\tau}\right\} = E\left\{\left\|\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right\|^2 - \frac{1}{2}\sum_{m=1}^{D^2-1}\sum_{j,k,l} f_{k,m,l}f_{k,m,j}\tau_l H_{k,k}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right)_j\right\}.$$

In the above expression, the term $\partial\boldsymbol{\beta}/\partial\omega$ is deterministic and piecewise constant in the case of GD compensation. Using the linearity of expectation operator, we obtain

$$\frac{\partial\boldsymbol{\beta}}{\partial\omega}\cdot\frac{dE\{\boldsymbol{\tau}\}}{dz} = \left\|\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right\|^2 - \frac{1}{2}\sum_{m=1}^{D^2-1}\sum_{j,k,l} f_{k,m,l}f_{k,m,j}E\{\tau_l\}H_{k,k}\left(\frac{\partial\boldsymbol{\beta}}{\partial\omega}\right)_j. \tag{49}$$

In (49), all of the terms have a common $\partial\boldsymbol{\beta}/\partial\omega$ dot product operation. Removing this $\partial\boldsymbol{\beta}/\partial\omega$ common factor yields:

$$\frac{\partial}{\partial z}E\{\boldsymbol{\tau}\} = \frac{\partial\boldsymbol{\beta}}{\partial\omega} - \mathbf{Q}\cdot E\{\boldsymbol{\tau}\} \tag{50}$$

where the elements of the $\left(D^2-1\right)\times\left(D^2-1\right)$ matrix $\mathbf{Q}$ are given by

$$\mathbf{Q}_{j,l} = \frac{1}{2}\sum_{m,k=1}^{D^2-1} f_{k,m,l}f_{k,m,j}H_{k,k}. \tag{51}$$

In this paper, we focus on the case of diagonal $\mathbf{H}$ for the coupling model considered. For a general covariance matrix $\mathbf{H}$, the derivations above are modified slightly and can be expressed in terms of its eigenvalue decomposition $\mathbf{H} = \mathbf{U}\mathbf{N}\mathbf{U}^T$ where $\mathbf{N}$ is the diagonal matrix of eigenvalues. The basis change from $\mathbf{H}$ to $\mathbf{N}$ does not change (41) because both $\boldsymbol{\tau}$ and $\partial\boldsymbol{\beta}/\partial\omega$ rotate simultaneously, and the inner product between them remains constant. On the other hand, $\mathbf{Q}$ in (50) can be replaced by $\mathbf{U}\mathbf{Q}\mathbf{U}^T$ while replacing $H_{k,k}$ in (51) by $N_{k,k}$.

C. Optimal GD Compensation Lengths

Assume a system that concatenates n different fibers types, each supporting D modes, with negligible mode coupling. Let $\boldsymbol{\tau}_i = (\boldsymbol{\tau}_{i,1}, \boldsymbol{\tau}_{i,2}, \ldots, \boldsymbol{\tau}_{i,D})$ and $L_{C,i}$ denote the GDs and length of the ith fiber for $1 \le i \le n$. Constraining the span length $\sum_{i=1}^{n} L_{C,i} = L_{amp}$, our goal is to find the optimal lengths $L_{C,i}$, $1 \le i \le n$ (which have $n-1$ degrees of freedom) for the given $\boldsymbol{\tau}_i$ that minimize the peak-to-peak GD spread $\max\left(\sum_{i=1}^{n} L_{C,i}\boldsymbol{\tau}_i\right) - \min\left(\sum_{i=1}^{n} L_{C,i}\boldsymbol{\tau}_i\right)$. We observe that if $\boldsymbol{\tau}_j$ is a linear combination of $\boldsymbol{\tau}_i$ for $1 \le i \le n$ and $i \ne j$, the minimum peak-to-peak GD spread is zero. In all other cases, a zero peak-to-peak GD spread cannot be obtained.

Determination of the optimal $L_{C,i}$ can be expressed as a linear program [41] with the following objective and constraints:

$$\text{minimize } \tau_{T,\max} - \tau_{T,\min}, \tag{52}$$

$$\text{subject to} \sum_{i=1}^{n} L_{C,i}\boldsymbol{\tau}_i - \tau_{T,\max}\cdot\mathbf{1} \le 0, \tag{53}$$

$$-\sum_{i=1}^{n} L_{C,i}\boldsymbol{\tau}_i + \tau_{T,\min}\cdot\mathbf{1} \le 0, \tag{54}$$

$$\sum_{i=1}^{n} L_{C,i} - L_{amp} = 0. \tag{55}$$

In the above, $\tau_{T,\max}$, $\tau_{T,\min}$ and $L_{C,i}$ are variables and $\mathbf{1}$ denotes a $D \times 1$ vector of ones. This linear program can be solved using a standard optimization package, e.g., [42].

REFERENCES

[1] R. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Labs Tech. J.*, vol. 14, no. 4, pp. 3–9, Feb. 2010.

[2] R. Ryf *et al.*, "Mode-division multiplexing over 96 km of few-mode fiber using coherent 6 × 6 MIMO processing," *J. Lightw. Technol.*, vol. 30, no. 4, pp. 521–531, Feb. 2012.

[3] S. O. Arık and J. M. Kahn, "Diversity-multiplexing tradeoff in mode-division multiplexing," *Opt. Lett.*, vol. 39, no. 11, pp. 3258–3261, Jun. 2014.

[4] K.-P. Ho and J. M. Kahn, "Mode coupling and its impact on spatially multiplexed systems," in *Optical Fiber Telecommunications VI*, I. P. Kaminow, T. Li, and A. E. Willner, Eds. New York, NY, USA: Elsevier, 2013.

[5] P. Winzer, "Making spatial multiplexing a reality," *Nature Photon.*, vol. 8, pp. 345–348, May 2014.

[6] S. O. Arik, K.-P. Ho, and J. M. Kahn, "Optical network scaling: Roles of spectral and spatial aggregation," *Opt. Exp.*, vol. 22, no. 24, pp. 29868–29887, Nov. 2014.

[7] S. O. Arik, D. Askarov, and J. M. Kahn, "Adaptive frequency-domain equalization in mode-division multiplexing systems", *J. Lightw. Technol.*, vol. 32, no. 10, pp. 1841–1852, May 2014.

[8] S. O. Arik, J. M. Kahn, and K.-P. Ho, "MIMO signal processing for mode-division multiplexing," *IEEE Signal Process. Mag.*, vol. 31, no. 2, pp. 25–34, Mar. 2014.

[9] S. O. Arik, D. Askarov, and J. M. Kahn, "Effect of mode coupling on signal processing complexity in mode-division multiplexing," *J. Lightw. Technol.*, vol. 31, no. 3, pp. 423–431, Feb. 2013.

[10] D. Peckham, Y. Sun, A. McCurdy, and R. Lingle, "Few-mode fiber technology for spatial multiplexing," in *Optical Fiber Telecommunications VI A*, I. P. Kaminow, T. Li, and A. E. Willner, Eds. New York, NY, USA: Elsevier, 2013.

[11] D. Marcuse, "The impulse response of an optical fiber with parabolic index profile," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1169–1174, Sep. 1973.

[12] P. Sillard, M. Astruc, and D. Molin, "Few-mode fibers for mode-division-multiplexed systems," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2824–2829, Aug. 2014.

[13] T. Mori *et al.*, "Few-mode fibers supporting more than two LP modes for mode-division-multiplexed transmission with MIMO DSP," *J. Lightw. Technol.*, vol. 32, no. 14, pp. 2468–2479, Jul. 2014.
[14] K.-P. Ho and J. M. Kahn, "Statistics of group delays in multimode fiber with strong mode coupling," *J. Lightw. Technol.* vol. 29, no. 21, pp. 3119–3128, Nov. 2011.
[15] M. Li *et al.*, "Low delay and large effective area few-mode fibers for mode-division multiplexing," presented at the *Opto-Electronics Communications Conf.*, Busan, Korea, 2012, Paper 5C3-2.
[16] T. Sakamoto, T. Mori, T. Yamamoto, and S. Tomita, "Differential mode delay managed transmission line for WDM-MIMO system using multi-step index fiber," *J. Lightw. Technol.*, vol. 30, no. 17, pp. 2783–2787, Sep. 2012.
[17] S. Randel *et al.* "Mode-multiplexed 6×20-GBd QPSK transmission over 1200-km DGD-compensated few-mode fiber" presented at the *Optical Fiber Communication Conf.*, Los Angeles, CA, USA, 2012, Paper PDP5C.5.
[18] N. Bai and G. Li, "Equalizer tap length requirement for mode group delay-compensated fiber link with weakly random mode coupling," *Opt. Exp.*, vol. 22, no. 4, pp. 4247–4255, Feb. 2014.
[19] S. D. Personick, "Time dispersion in dielectric waveguides," *Bell Syst. Tech. J.*, vol. 50, no. 3, pp. 843–859, Mar. 1971.
[20] M. J. Li and D. A. Nolan, "Fiber spin-profile designs for producing fibers with low polarization mode dispersion," *Opt. Lett.*, vol. 23, pp. 1659–1661, Nov. 1998.
[21] K.-P. Ho and J. M. Kahn, "Linear propagation effects in mode-division multiplexing systems," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 614–628, Feb. 2014.
[22] L. Palmieri, "Modal dispersion properties of few-mode spun fibers," presented at the *Optical Fiber Communication Conf.*, Los Angeles, CA, USA, 2015, Paper Tu2D.4.
[23] S. O. Arık, D. Askarov, and J. M. Kahn, "MIMO DSP complexity in mode-division multiplexing systems," presented at the *Optical Fiber Communication Conf.*, Los Angeles, CA, USA, 2015, Paper Th1D.1.
[24] S. Fan and J. M. Kahn, "Principal modes in multi-mode waveguides," *Opt. Lett.*, vol. 30, no. 2, pp. 135–137, Jan. 2005.
[25] C. D. Poole and R. E. Wagner, "Phenomenological approach to polarization dispersion in long single-mode fibers," *Electron. Lett.*, vol. 22, pp. 1029–1030, Sep. 1986.
[26] E. Ip and J. M. Kahn, "Fiber impairment compensation using coherent detection and digital signal processing," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 502–519, Feb. 2010.
[27] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers," *J. Lightw. Technol.*, vol. 9, no. 11, pp. 1439–1456, Nov. 1991.
[28] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization-mode dispersion in optical fibers," *Proc. Nat. Acad. Sci.*, vol. 97, no. 9, pp. 4541–4550, Apr. 2000.
[29] C. Antonelli, A. Mecozzi, M. Shtaif, and P. Winzer, "Stokes-space analysis of modal dispersion in fibers with multiple mode transmission," *Opt. Exp.*, vol. 20., no. 11, pp. 11718–11733, May 2012.
[30] C. Antonelli, A. Mecozzi, M. Shtaif, and P. J. Winzer, "Random coupling between groups of degenerate fiber modes in mode multiplexed transmission," *Opt. Exp.*, vol. 21, no. 8, pp. 9484–9489, Apr. 2013.
[31] C. Antonelli, A. Mecozzi, and M. Shtaif, "The delay spread in fibers for SDM transmission: Dependence on fiber parameters and perturbations," *Opt. Exp.*, vol. 23, no. 3, pp. 2196–2202, Feb. 2015.
[32] C. W. Gardiner, *Handbook of Stochastic Methods*, 2 ed. New York, NY, USA: Springer-Verlag, 1985.
[33] R. Ryf *et al.*, "12 × 12 MIMO transmission over 130-km few-mode fiber," presented at the *Frontiers Optics Conf.*, Rochester, NY, USA, Oct. 2012, Paper FW6C.4.
[34] K.-P. Ho and J. M. Kahn, "Mode-dependent loss and gain: Statistics and effect on mode-division multiplexing," *Opt. Exp.*, vol. 19, no. 17, pp. 16612–16635, Aug. 2011.
[35] 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, Dec. 2012.
[36] A. Mecozzi, C. Antonelli, and M. Shtaif, "Nonlinear propagation in multimode fibers in the strong coupling regime," *Opt. Exp.*, vol. 20, no. 11, pp. 11673–11678, May 2012.
[37] A. Mecozzi, C. Antonelli, and M. Shtaif, "Coupled Manakov equations in multimode fibers with strongly coupled groups of modes," *Opt. Exp.*, vol. 20, no. 21, pp. 23436–23441, Oct. 2012.
[38] G. Kimura, "The Bloch vectors for *N*-levels systems," *Phys. Lett. A*, vol. 314, no. 5, pp. 339–349, Aug. 2003.
[39] V. S. Varadarajan, *Lie Groups, Lie Albegras, and Their Representations*. New York, NY, USA: Springer-Verlag, 1984.
[40] K.-P. Ho, "Non-Gaussian statistics of soliton timing jitter induced by amplifier noise," *Opt. Lett.*, vol. 28, no. 22, pp. 2165–2167, Nov. 2003.
[41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[42] M. Grant, S. Boyd, and Y. Ye, *Disciplined Convex Programming, in Global Optimization: From Theory to Implementation*, L. Liberti and N. Maculan Eds. New York, NY, USA: Springer-Verlag, 2006, pp. 155–210. Available: http://www.stanford.edu/boyd/cvx/

Sercan Ö. Arık received the B.S. degree from Bilkent University, Ankara, Turkey, and the M.S. degree from Stanford University, Stanford, CA, USA, in electrical engineering in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree at Stanford University. He was with EPFL Signal Processing Labs, Lausanne, Switzerland, in summer 2010, with Google, Mountain View, CA, in summer 2012, and with Mitsubishi Electric Research Labs, Cambridge, MA, USA, in summer 2013. His main current research interests include multimode optical communications, advanced modulation, coding and digital signal processing techniques for fiber-optic systems, and optical signal processing.

Keang-Po Ho (S'91–M'95–SM'03) received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1991, and the M.S. and Ph.D. degrees from the University of California at Berkeley, Berkeley, CA, USA, in 1993 and 1995, respectively, all in electrical engineering. He has been with SiBEAM (acquired first by Silicon Image and later by Lattice Semiconductor) since 2006 as the Director of Engineering for the Baseband Algorithm Group, responsible for the invention of WirelessHD video-area networking, design and implementation of both WirelessHD and WiGig wireless systems using 60-GHz millimeter-wave antenna array, 3.8-Gb/s OFDM modem, and adaptive beam forming. He was the Chief Technology Officer and Cofounder of StrataLight Communications (acquired by OpNext by about US$170M). He had also been with IBM T. J. Watson Research Center, Bellcore (Telcordia), the Chinese University of Hong Kong, and National Taiwan University. Over the years, his major research fields include high-speed digital communication systems via optical fiber, copper wire, or radio frequency. He authored more than 200 journal and conference articles, 15 issued U.S. patents, several book chapters, and the book *Phase-Modulated Optical Communication Systems* (New York, NY, USA: Springer-Verlag, 2005).

Joseph M. Kahn (M'90–SM'98–F'00) received the A.B., M.A., and Ph.D. degrees in physics from the University of California at Berkeley (U.C. Berkeley), Berkeley, CA, USA, in 1981, 1983, and 1986, respectively. From 1987 to 1990, he was with AT&T Bell Laboratories, Crawford Hill Laboratory, Holmdel, NJ, USA. He demonstrated multigigabit per second coherent optical fiber transmission systems, setting world records for receiver sensitivity. From 1990 to 2003, he was on the Faculty of the Department of Electrical Engineering and Computer Sciences, U.C. Berkeley, performing research on optical and wireless communications. Since 2003, he has been a Professor of electrical engineering at Stanford University, Stanford, where he heads the Optical Communications Group. His current research interests include fiber-based imaging, spatial multiplexing, rate-adaptive and spectrally efficient modulation and coding methods, coherent detection and associated digital signal processing algorithms, digital compensation of fiber nonlinearity, and free-space systems. He received the National Science Foundation Presidential Young Investigator Award in 1991. From 1993 to 2000, he was a Technical Editor of the IEEE PERSONAL COMMUNICATIONS MAGAZINE. Since 2009, he has been an Associate Editor of the IEEE/OSA JOURNAL OF OPTICAL COMMUNICATIONS AND NETWORKING. In 2000, he helped to found StrataLight Communications, where he served as the Chief Scientist from 2000 to 2003. StrataLight was acquired by Opnext, Inc., in 2009.