Group Delay Management and Multiinput Multioutput Signal Processing in Mode-Division Multiplexing Systems

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

(Invited Paper)

***Abstract*—Multi-input multi-output (MIMO) digital signal processing (DSP) for mode-division multiplexing (MDM) may have high complexity, owing to a plurality of modes and a potentially long group delay (GD) spread in multimode fiber (MMF). This paper addresses the management of GD in MMF and its implications for the complexity and performance of MIMO DSP. First, we review the generalized Jones and Stokes representations for modeling propagation in MMF, and describe key GD properties derived using the two representations. Then, we describe three approaches for GD management: 1) optimized fiber design, 2) mode coupling, and 3) GD compensation. For approach 1), we explain design principles for minimizing the GD spread. We review experimental results to date, showing that fabrication nonidealities significantly increase the GD spread, and this approach alone may not achieve sufficiently low GD spread. For approach 2), we describe mechanisms for inducing intragroup and intergroup coupling. We describe mode scrambler designs based on photonic lanterns or long-period fiber gratings, both of which can ensure strong intergroup coupling. For approach 3), we review GD-compensated system design principles and show that GD compensation is only partially effective in the presence of random intragroup or intergroup coupling. Finally, we provide an overview of adaptive MIMO frequency-domain equalization algorithms. Considering tradeoffs between complexity, performance, and adaptation time, we show that the GD spread is a key factor determining the feasibility of MIMO DSP, and its feasibility requires judicious GD management.**

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

I. Introduction

WORLDWIDE data traffic has been growing exponentially over the past few decades, and the trend is expected to continue in the foreseeable future [1]. In long-haul systems, traffic growth to date has been accommodated using single-mode fiber (SMF). Contemporary SMF systems multiplex information in all available physical degrees of freedom. The throughput per SMF, after decades of exponential growth, is approaching estimated information-theoretical limits [1]–[4]. These circumstances have motivated research into spatial multiplexing schemes for increasing the transmission capacity per fiber.

The orthogonal modes of multimode fiber (MMF) are a promising option for multiplexing to increase per-fiber capacity. Using multi-input multi-output (MIMO) transmission techniques inspired by wireless communications, mode-division multiplexing (MDM) in MMFs can achieve a capacity per fiber that is ideally proportional to the number of propagating modes, denoted here by D [5]–[10]. Compared to transmission in multiple-fiber bundles or multi-core fibers (MCFs), MDM in MMFs yields a much higher spatial information density [11]. This high spatial information density enables highly compact design of amplifiers, switches and other inline components, which aids in the economical and energy-efficient scaling of optical networks [7], [11].

However, the high spatial information density of MDM in MMF comes at the expense of crosstalk caused by mode coupling and distortion caused by modal dispersion (MD). An MDM receiver must employ adaptive MIMO equalization to compensate for these, while tracking fast changes in the MIMO channel [13], [14]. A MIMO equalizer must have spatial size sufficient to compensate all pairwise crosstalk effects, requiring a size $D \times D$ in the worst case. The equalizer must have temporal memory sufficient to span the duration of the MIMO channel. In long-haul systems, to an excellent approximation in the regimes of interest, the MIMO channel duration is the peak-to-peak (*p–p*) coupled group delay (GD) spread τ_{p-p} [15]–[18].

SMF systems ($D = 2$) already employ adaptive 2×2 MIMO equalization to compensate for time-varying polarization crosstalk and polarization mode dispersion (PMD), which are analogues of mode coupling and MD [19]–[21]. In SMF, the PMD-induced GD spread is low, so the 2×2 MIMO equalization is often implemented in the time domain. In long-haul SMF systems, where the 2×2 MIMO channel can change on the time scale of tens of microseconds [22], adaptive MIMO equalization can achieve sufficiently fast channel tracking [21], [23]. While satisfying performance and adaptation speed requirements, this 2×2 MIMO digital signal processing (DSP) consumes about 5–15% of the total power in a contemporary long-haul link [24].

Scaling up to MMF ($D > 2$) poses significant challenges for adaptive MIMO equalization [12]–[14], [25]. If non-negligible crosstalk occurs only between degenerate modes, the

Manuscript received July 19, 2015; revised January 2, 2016 and February 3, 2016; accepted February 13, 2016. Date of publication February 17, 2016; date of current version April 30, 2016. This work was supported by a Google Research Award and by Corning, Inc.

S. Ö. Arık and J. M. Kahn are with the 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, 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.2016.2530978

spatio-temporal sparsity of the channel may be exploited by employing separate MIMO equalizers for the different mode groups, each with a size and temporal duration determined by the respective group [26], [27]. In experiments to date, however, long-haul MDM systems have exhibited non-negligible crosstalk between all mode pairs [28]–[31], requiring a full $D \times D$ MIMO equalizer with length spanning the *p–p* GD spread τ_{p-p}. The GD spread in MMF can be orders of magnitude higher than in SMF systems [12]. The ability of a MIMO equalizer to track dynamic MIMO channels is also an important consideration. MIMO channel variations in MMFs are expected to be at least as fast as those in SMFs [32], [33]. Given the large $D \times D$ equalizer size and the long GD spread, achieving sufficiently fast channel tracking may be challenging [13]. As a result of the above, MIMO DSP is likely to be a major contributor to cost and power consumption in MDM systems, and might limit the feasibility of MDM technology.

Notable MDM experiments to date [28], [30], [34]–[37], have focused on demonstrating high-throughput transmission and integration of multimode optical components. They implemented MIMO DSP using offline processing, without concern for computational complexity or dynamic channel tracking. To date, real-time MIMO DSP for MDM has been demonstrated only over relatively short lengths ($\sim$60 km) of coupled-core MCFs [38], which have far lower uncoupled GD spread per unit length than MMFs [39], [40]. The practicality of real-time MIMO DSP in long-haul MDM systems remains a critical open question.

This paper provides an overview of GD management techniques in MMFs and their implications for MIMO DSP in MDM systems, addressing underlying principles, experimental demonstrations, challenges, and future directions. The main contributions of this paper are as follows. Section II reviews approaches for modeling fiber propagation and the GD operator in both generalized Jones and Stokes representations. Major results obtained using the two approaches in the regimes of practical interest are described and compared. Section III focuses on methods for reducing system end-to-end GD spread, including fiber design optimization, mode coupling, and GD compensation. Section III-A explains fiber GD characterization techniques, which helps relate GD modeling approaches to experimentally realized fiber designs. Section III-B presents the GD characteristics of fibers fabricated to date, demonstrating the significant gap between theoretical and experimental GD properties. Section III-C explains how the strong mode coupling regime may be realized using mode scramblers. Scramblers using long-period fiber gratings (LPFGs) are reviewed, and scramblers using photonic lanterns are proposed for the first time. Section III-D explains GD spread reduction using GD compensation. Experimental demonstrations are reviewed and, using the GD modeling techniques from Section II, the impact of mode coupling on GD compensation is explained. Section IV describes MIMO DSP architectures and algorithms, and the impact of GD spread on their performance and complexity. MIMO DSP algorithms demonstrated to date are categorized and compared in terms of their complexity and adaptation time. Sections V and VI present discussion and conclusion, respectively.

II. GD Modeling

Understanding GD management and MIMO DSP requires detailed characterization of the GD operator and, ultimately, the statistics of the *p–p* GD spread τ_{p-p}. The *p–p* GD spread provides a good first-order approximation to the duration of the intensity impulse response of the channel [15] and the duration of the impulse response of the required equalizer. Second- and higher-order MD effects increase the impulse response duration slightly in the regime when the MD (coherence) bandwidth is much smaller than the signal bandwidth, i.e., the *p–p* GD spread spans many symbol intervals [15]. In long-haul MDM systems, this regime is highly undesirable because of the high associated MIMO DSP complexity [12].

To simplify the discussion, we neglect mode-averaged gain and phase effects, mode-dependent gains/losses (MDL) and mode-dependent chromatic dispersion (CD), in which case, linear propagation in a fiber can be described by a GD operator [27], whose eigenvectors are the principal modes (PMs). In the case of SMF ($D = 2$), representations of the GD operator in Jones [41], [42] and Stokes [42], [43] 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 [41]. The GDs of the PSPs are equal and opposite, so the differential GD (DGD) is sufficient to represent them. 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 also fully represent the GD operator [42].

In the case of MMF ($D > 2$), both the generalized Jones [44] and Stokes [45] representations of the GD operators may be used to study GD properties. These representations are reviewed in the following two subsections.

A. Generalized Jones Representation

In generalized Jones representation, end-to-end propagation is described by a $D \times D$ zero-trace Hermitian matrix $\mathbf{M}(\Omega)$, which multiplies *D* complex baseband modal envelopes at each frequency Ω. Given $\mathbf{M}(\Omega)$, the GD operator is defined as

$$\mathbf{G} = j\frac{\partial \mathbf{M}(\Omega)}{\partial \Omega}\mathbf{M}(\Omega)^{H}, \tag{1}$$

where ^{H} denotes Hermitian transpose. The eigenvectors of $\mathbf{G}$ are the PMs [46], and the corresponding eigenvalues are their GDs, which may be referred to as the coupled GDs. For the coupled GDs $(\tau_1^{(t)}, \ldots, \tau_D^{(t)})$, we assume the ordering $\tau_1^{(t)} \leq \tau_2^{(t)} \leq \cdots \leq \tau_D^{(t)}$, so the *p–p* GD spread is $\tau_{p-p} = \tau_D^{(t)} - \tau_1^{(t)}$.

Generalizing from SMF [20], $\mathbf{M}(\Omega)$ is often described using a multi-section model as

$$\mathbf{M}(\Omega) = \prod_{k=1}^{K_{\mathrm{amp}}} \prod_{l=1}^{K_{\mathrm{sec}}} \mathbf{V}^{(k,l)} \mathrm{diag}\left[e^{-j\Omega\tau_1 L_{\mathrm{sec}}} \cdots e^{-j\Omega\tau_D L_{\mathrm{sec}}}\right] \mathbf{U}^{(k,l)H}. \tag{2}$$

The multi-section model describes an end-to-end system as a concatenation of fiber sections, each with a length roughly equal to the correlation length for propagating fields [27], [44]. The multi-section model (2) assumes that the system is comprised of $K_{\rm amp}$ identical fiber spans, each composed of $K_{\rm sec}$ identical sections of length $L_{\rm sec}$, such that the total system length is $L_{\rm tot} = K_{\rm amp}K_{\rm sec}L_{\rm sec}$. Each diagonal matrix in (2) describes uncoupled propagation in the corresponding section.[1] In each diagonal matrix, the delays τ_i, $i = 1, \ldots, D$, are the uncoupled modal GDs per unit length, with $\sum_{i=1}^{D} \tau_i = 0$ and r.m.s. value σ_τ. If GD compensation is employed (as in Section III-D), the τ_i values depend on the fiber type in the corresponding section.

In (2), mode coupling is modeled by inserting matrices $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ between the diagonal matrices representing uncoupled propagation. More specifically, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are frequency-independent unitary matrices representing mode coupling in the *l*th section of the *k*th span. In the absence of mode coupling, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are $D \times D$ identity matrices, and the end-to-end GD operator is $\mathbf{G} = K_{\rm sec}K_{\rm amp}L_{\rm sec}{\rm diag}[\,\tau_1 \ \ldots \ \tau_D\,]$, so the coupled GDs and GD spread increase linearly with propagation distance. In the presence of mode coupling, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ may be deterministic, e.g., to model intentional perturbations or mode scrambler devices. Alternatively, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ may be random, with statistics determined by the random perturbations, and the statistics of the eigenvalues of $\mathbf{G}$ and the corresponding GDs may be studied using random matrix theory [6], [27], [44].

The generalized Jones representation yields analytical GD statistics in the regime of strong mode coupling induced by random perturbations [27], [44], [47]. In this regime, mode coupling occurs with approximately equal strength between all modes, such that $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are fully random unitary matrices, and the number of independent sections is large, i.e., $K_{\rm sec}K_{\rm amp} \gg 1$ [44]. The coupled r.m.s. GD spread is

$$\sigma_{\rm gd} = \sqrt{K_{\rm sec}K_{\rm amp}}L_{\rm sec}\sigma_\tau = \sqrt{L_{\rm tot}L_{\rm sec}}\sigma_\tau. \qquad (3)$$

The coupled GDs $(\tau_1^{(t)}, \ldots, \tau_D^{(t)})$ are random variables. For a given D, their statistics depend on a single parameter, which may be taken as the coupled r.m.s. GD $\sigma_{\rm gd}$ [44], the mean of $|\tau_i^{(t)}|$, the *p–p* GD spread $\tau_{p-p} = \tau_D^{(t)} - \tau_1^{(t)}$ [47], or a MD (coherence) bandwidth [15]. In the strong-coupling regime, for all *D*, the *p–p* GD spread $\tau_{p-p} = \tau_D^{(t)} - \tau_1^{(t)}$ is roughly $4\sigma_{\rm gd}$ to $5\sigma_{\rm gd}$ [14], [47] at all frequencies. (The channel impulse response duration may exceed the *p–p* GD spread slightly because of frequency-dependent second- and higher-order effects [15], but it too depends solely on the coupled r.m.s. GD $\sigma_{\rm gd}$.) Thus, the system end-to-end *p–p* GD spread scales as $\sqrt{L_{\rm tot}}$, the square-root of propagation distance. Moreover, the end-to-end *p–p* GD spread scales as $\sqrt{L_{\rm sec}}$, so it is decreased by decreasing the section length, which corresponds to increasing the coupling strength.

[1]Throughout this paper, uncoupled propagation is described in a basis of ideal vector modes, which are propagation-invariant field patterns computed without approximation for an unperturbed fiber, and the $\tau_{\rm i}$ represent uncoupled GDs per unit length for the vector modes. In Section III-B, to facilitate comparison with experimental literature, we present peak-to-peak GD spreads for LP modes derived from the vector modes.

As explained in Section III-C, many random perturbations strongly couple modes having nearly equal propagation constants (intragroup coupling), but weakly couple modes having highly unequal propagation constants (intergroup coupling) [27]. When only intragroup coupling occurs, e.g., in short-haul systems, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are block unitary matrices that depend on the group structure. For example, for $D = 12$ modes in groups of 2, 4 and 6 modes, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ incorporate random unitary random submatrices of size 2×2, 4×4 and 6×6. When $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are block unitary, $\mathbf{M}(\Omega)$ becomes block unitary and $\mathbf{G}$ becomes block Hermitian. In this regime, the MIMO equalizer can be simplified to a block unitary structure to exploit spatio-temporal sparsity [26].

Depending on the strength of intragroup coupling within each mode group, the section lengths may be different for different mode groups. The GD statistics for a system with only intragroup coupling can be studied analytically using the generalized Jones representation. Each block of $\mathbf{G}$ is a combination of two random matrices: a zero-trace Gaussian unitary ensemble (GUE) and a group-mean matrix. The GUE describes intragroup coupling. The statistics of its eigenvalues depend on the r.m.s. GD spread of the mode group and the number of modes in the group. The group-mean matrix is proportional to the group's mean GD times an identity matrix, and serves to shift the GUE eigenvalues by the mean GD. Since there is no coupling between mode groups, differences between the means of the mode groups accumulate linearly with propagation distance. The overall probability density of the coupled GDs can be found by composing the probability densities for the different groups with appropriately shifted means.

The generalized Jones representation can provide the coupled GDs directly from the eigenvalues of $\mathbf{G}$. Although concatenation rules can be found for multiple fiber sections [27], it is not convenient to analytically model the evolution of $\mathbf{G}$ in generalized Jones space. As explained next, the generalized Stokes representation is more convenient for modeling the continuous evolution of **G,** because random mode coupling can be represented as an additive noise, and analytical methods for additive stochastic processes are well-developed.

B. Generalized Stokes Representation

In generalized Stokes representation, the GD operator and propagation are described using a set of basis matrices and corresponding vector operators. The GD operator (1) is a $D \times D$ zero-trace Hermitian matrix, and a convenient basis set is a set of trace-orthogonal matrices $\mathbf{\Lambda}_i$ $(1 \le i \le D^2 - 1)$, e.g., the generalized Pauli matrices [45]. The representation of the GD operator in generalized Stokes space is defined by a set of $D^2 - 1$ coefficients multiplying the basis matrices. The coupled GD can be studied through the propagation of the generalized Stokes vectors [45], [48]. Generalizing from SMF, their propagation may be studied using a stochastic differential equation

$$\frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\partial \boldsymbol{\beta}}{\partial \Omega} + \boldsymbol{\beta} \times \boldsymbol{\tau}, \qquad (4)$$

where $\boldsymbol{\beta}$ is a $(D^2 - 1) \times 1$ birefringence vector, $\boldsymbol{\tau}$ is a $(D^2 - 1) \times 1$ dispersion vector, z is propagation distance and

Ω is angular frequency. The cross product $\times$ is defined for the generalized Stokes space using structure constants [45], [48]. The propagation equation (4) is simpler than the corresponding concatenation rule in generalized Jones representation [27]. Generalizing the linear birefringence model from SMF [43], the z dependence of the birefringence vector can be expressed as

$$\boldsymbol{\beta}(z) = \frac{\partial \boldsymbol{\beta}(z)}{\partial \Omega} \cdot \Omega + \boldsymbol{g}(z). \tag{5}$$

The $\partial\boldsymbol{\beta}/d\Omega$ term in (5) represents the deterministic GD, such that $\mathrm{diag}(\tau_1, \ldots, \tau_D) = (\sum_{i=1}^{D^2-1} (\partial\boldsymbol{\beta}/d\Omega)_i \boldsymbol{\Lambda}_i)/D$, where $(\tau_1, \ldots, \tau_D)$ are the uncoupled GDs per unit length, which also appear in the diagonal term of (2). The modulus of $\partial\boldsymbol{\beta}/d\Omega$, $\sqrt{\|\partial\boldsymbol{\beta}/d\Omega\|^2}$, gives the uncoupled r.m.s. GD per unit length σ_τ, with a scale factor of D [48]. If GD compensation is employed (see Section III-D), $\partial\boldsymbol{\beta}/d\Omega$ depends on the fiber type at the corresponding location.

Random mode coupling is represented by the (D^2-1)-dimensional additive noise vector $\mathbf{g}(z)$ in (5). In SMF, the three-dimensional random perturbation $\mathbf{g}(z)$ is assumed to be distributed as $N(\mathbf{0}, (1/h)\mathbf{I})$, where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multi-dimensional Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and $\mathbf{W}(l) = \int_0^l \sqrt{h}\mathbf{g}(z)dz$ is a 3-D standard Wiener process (Brownian motion), such that $\mathbf{W}(0) = \mathbf{0}$, and $\mathbf{W}(z) \sim N(\mathbf{0}, z\mathbf{I})$. 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 $z \gg h$, while weak coupling corresponds to $z \ll h$ [43].

As explained in detail in Section III-C, intragroup and intergroup coupling typically have different coupling strengths. To account for this while generalizing the coupling model from SMF in the simplest possible way, we define two different coupling lengths h_{intra} and h_{inter} for intragroup and intergroup coupling, respectively [48]. A similar approach is used to distinguish the strengths of polarization coupling and inter-core spatial coupling in MCFs [49]. For a D-mode fiber, we assume the (D^2-1)-dimensional $\mathbf{g}(z)$ is distributed as $\mathbf{g}(z) \sim N(\mathbf{0}, \mathbf{H})$, where $\mathbf{H}$ is a diagonal covariance matrix, and $\mathbf{W}(l) = \int_0^l \mathbf{H}^{-0.5}\mathbf{g}(z)dz$ is a standard (D^2-1)-dimensional Wiener process, such that $\mathbf{W}(0) = \mathbf{0}$, and $\mathbf{W}(z) \sim N(\mathbf{0}, z\mathbf{H})$. The diagonal elements of $\mathbf{H}$ are assumed to be either $1/h_{\mathrm{intra}}$ or $1/h_{\mathrm{inter}}$, depending on the choice of the basis matrices. [2]

The GD statistics are governed by the stochastic evolution of the dispersion vector $\boldsymbol{\tau}$. The PMs and coupled GDs are given by the eigenvectors and eigenvalues of $\mathbf{G} = (\sum_{i=1}^{D^2-1} \tau_i \boldsymbol{\Lambda}_i)/D$. In the absence of random perturbations, $\boldsymbol{\tau}$ is deterministic and equals $z\partial\boldsymbol{\beta}/\partial\Omega$, so that $\boldsymbol{\tau}/z = \boldsymbol{\beta}/\omega = \partial\boldsymbol{\beta}/\partial\Omega$ with vectors $\boldsymbol{\tau}$, $\boldsymbol{\beta}$ and $\partial\boldsymbol{\beta}/\partial\Omega$ aligned. The coupled GDs are equal to the uncoupled GDs, which accumulate linearly with propagation distance. In the presence of random perturbations, the vectors $\boldsymbol{\beta}$ and $\partial\boldsymbol{\beta}/\partial\Omega$ become misaligned from each other, depending on the coupling strength. The direction and modulus of $\boldsymbol{\tau}$ changes stochastically at each increment of propagation distance. The coupled GDs become random variables.

The generalized Stokes representation enables analytical derivation of the average r.m.s. GD in various coupling regimes. The r.m.s. GD for the end-to-end system can be conveniently represented as $\sqrt{\boldsymbol{\tau}\cdot\boldsymbol{\tau}}/D = \sqrt{\|\boldsymbol{\tau}\|^2}/D$ [45], [48], i.e., it is independent of the direction of $\boldsymbol{\tau}$, and depends only on its modulus $\|\boldsymbol{\tau}\|$. The evolution of the probability density of $\boldsymbol{\tau}$ with z is described by Fokker–Planck equations, and the expected value of a function of $\boldsymbol{\tau}$ can be described by a set of deterministic ordinary differential equations [48]. The evolution of $E\{\|\boldsymbol{\tau}\|^2\}$ is given by [48], [49]

$$\frac{\partial}{\partial z} E\left\{\|\boldsymbol{\tau}\|^2\right\} = E\left\{2\frac{\partial\boldsymbol{\beta}}{\partial\Omega}\cdot\boldsymbol{\tau}\right\} = 2\frac{\partial\boldsymbol{\beta}}{\partial\Omega}\cdot E\{\boldsymbol{\tau}\}, \tag{6}$$

and the evolution of $E\{\boldsymbol{\tau}\}$ is given by [48]

$$\frac{\partial}{\partial z} E\{\boldsymbol{\tau}\} = \frac{\partial\boldsymbol{\beta}}{\partial\Omega} - \mathbf{Q}\cdot E\{\boldsymbol{\tau}\}. \tag{7}$$

In (7), the elements of the $(D^2-1)\times(D^2-1)$ matrix $\mathbf{Q}$ depend on the mode coupling characteristics. For the model discussed here, $\mathbf{Q}$ can be expressed in terms of h_{intra} and h_{inter} [48].

When $\partial\boldsymbol{\beta}/\partial\Omega$ is constant, corresponding to a system using one fiber type, the differential equation (7) is linear in $E\{\boldsymbol{\tau}\}$. Assuming an initial condition $E\{\boldsymbol{\tau}\} = \mathbf{0}$ at $z = 0$, integration of (7) yields:

$$E\{\boldsymbol{\tau}\} = \mathbf{Q}^{-1}\left(\mathbf{I} - \exp(-z\mathbf{Q})\right)\frac{\partial\boldsymbol{\beta}}{\partial\Omega}. \tag{8}$$

Inserting (8) into (6) and assuming an initial condition $E\{\|\boldsymbol{\tau}\|^2\} = 0$ at $z = 0$, integration of (6) yields:

$$E\left\{\|\boldsymbol{\tau}\|^2\right\} = 2\frac{\partial\boldsymbol{\beta}}{\partial\Omega}\cdot\left(\mathbf{Q}^{-1}z - \mathbf{Q}^{-2}\left(\mathbf{I} - \exp(-z\mathbf{Q})\right)\right)\frac{\partial\boldsymbol{\beta}}{\partial\Omega}. \tag{9}$$

Fig. 1 shows the evolution of the r.m.s. GD for various values of h_{intra} and h_{inter}. The r.m.s. GD accumulates linearly with distance z when z is much shorter than h_{intra} and h_{inter}. The r.m.s. GD accumulates with the square-root of distance z when z becomes much longer than h_{intra} and h_{inter}. When z is similar to h_{intra} and h_{inter}, the r.m.s. GD accumulates with a power of z lying between 1/2 and 1.

Although typically $h_{\mathrm{inter}} > h_{\mathrm{intra}}$, the assumption of $h_{\mathrm{inter}} = h_{\mathrm{intra}} = h$ allows modeling of mode coupling with a single parameter. In this case, with $\mathbf{H} = \mathbf{Q} = (1/h)\mathbf{I}$, the mean dispersion vector (8) simplifies to

$$E\{\boldsymbol{\tau}\} = h\left(1 - e^{-z/h}\right)\frac{\partial\boldsymbol{\beta}}{\partial\Omega}, \tag{10}$$

and the r.m.s. GD simplifies to [43]

$$E\left\{\|\boldsymbol{\tau}\|^2\right\} = 2\left\|\frac{\partial\boldsymbol{\beta}}{\partial\Omega}\right\|^2\left(hz - h^2\left(1 - e^{-z/h}\right)\right). \tag{11}$$

In the weak-coupling regime with $h \gg z$, we have $h^2 e^{-z/h} \approx h^2 - zh + z^2/2$. The r.m.s. GD is $\sqrt{E\{\|\boldsymbol{\tau}\|^2\}/D^2} \approx$

[2] The mode coupling model described here can be generalized to a broad range of scenarios by appropriate choice of $\mathbf{H}$. The intragroup and intergroup coupling strengths depend on the type of MMF and the perturbation [6], [49] (see Section III-C for more details). To include different coupling strength parameters in the model, different diagonal elements can be chosen for $\mathbf{H}$. It is also possible to include dependence between the elements in $\mathbf{g}(z)$, in which case, $\mathbf{H}$ will have non-diagonal elements, but will remain positive-semidefinite [47].

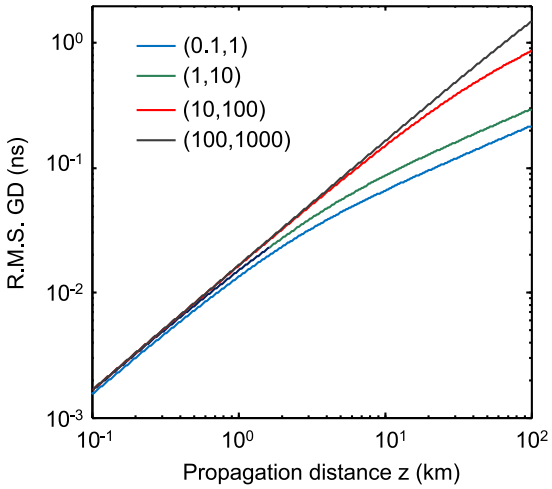

Fig. 1. Root-mean-square GD as a function of propagation distance z for $D = 12$ modes in a 100-km fiber span employing GIGDC fiber. The legends describe the intra- and inter-group coupling lengths $(h_{\mathrm{intra}}, h_{\mathrm{inter}})$ in kilometers.

$\sqrt{\|\partial\boldsymbol{\beta}/\partial\Omega\|^2/D^2 z}$, i.e., it is proportional to the r.m.s. uncoupled GD per unit length, and scales linearly with propagation distance z [43]. In the strong-coupling regime with $h \ll z$, we have $hz + h^2 e^{-z/h} - h^2 \approx hz$. The r.m.s. GD is $\sqrt{E\{\|\boldsymbol{\tau}\|^2\}/D^2} \approx \sqrt{2\|\partial\boldsymbol{\beta}/\partial\Omega\|^2 hz/D^2}$, i.e., it is proportional to r.m.s. uncoupled GD per unit length and scales with $\sqrt{hz}$ [43]. The ratio between the r.m.s. GD spreads in the strong- and weak-coupling regimes is $\sqrt{2h/z}$.

Finally, we can use the above results to relate the generalized Jones and Stokes representations. The coupling strength in the Jones representation of Section II-A is determined by the section length L_{sec} and the ratio between the r.m.s. GD spreads in the strong- and weak-coupling regimes is $\sqrt{L_{\mathrm{sec}}/z}$. Equating the r.m.s. GD spread reduction ratios for the two representations, the section length in the multi-section generalized Jones model is twice the coupling length in the generalized Stokes model with $L_{\mathrm{sec}} = 2h$. With a proper choice of parameters, the generalized Jones and Stokes representations are mutually consistent in the strong-coupling regime.

III. GD Management

A. GD Characterization

Accurate characterization of GDs is important for the design and analysis of MDM systems.

As stated in Section II-A, the uncoupled GDs $(\tau_1, \ldots, \tau_D)$ are the GDs per unit length of the exact vector modes of an unperturbed fiber. These GDs of vector modes may be difficult to characterize experimentally. Vector modes generally have spatially varying polarizations and small, but nonzero, longitudinal components, making them difficult to generate and detect. Vector modes have been generated and detected using computer-generated holograms with birefringent optics [51], [52]. Intragroup mode coupling complicates using vector modes in GD characterization. Vector modes in the same mode group may have different GDs, but have nearly equal propagation constants, so they couple significantly after short distances (see Section III-C), and impulse response measurements may yield distorted pulses [53], [54].

GDs are typically easier to characterize using linearly polarized (LP) "modes" that are linear combinations of nearly degenerate vector modes [55]. For example, $\mathrm{LP}_{11,\mathrm{a}}$ and $\mathrm{LP}_{11,\mathrm{b}}$ in x and y polarizations are formed from the vector modes TM_{01}, TE_{01}, $\mathrm{HE}_{21,\mathrm{a}}$ and $\mathrm{HE}_{21,\mathrm{b}}$. ($\mathrm{LP}_{11,\mathrm{a}}$ and $\mathrm{LP}_{11,\mathrm{b}}$ in the x polarization are $\mathrm{TM}_{01} - \mathrm{HE}_{21,\mathrm{a}}$ and $\mathrm{TE}_{01} + \mathrm{HE}_{21,\mathrm{b}}$, respectively, while $\mathrm{LP}_{11,\mathrm{a}}$ and $\mathrm{LP}_{11,\mathrm{b}}$ in the y polarization are $\mathrm{HE}_{21,\mathrm{b}} - \mathrm{TE}_{01}$ and $\mathrm{TM}_{01} + \mathrm{HE}_{21,\mathrm{a}}$, respectively). The LP modes have spatially invariant polarizations and zero longitudinal components, simplifying their generation and detection. LP modes have been generated and detected using various methods, e.g., from a LP fundamental mode using phase plates with profile matching the desired mode [28]. In the absence of intragroup mode coupling, LP modes are not truly propagation-invariant, due to slight differences between the propagation constants of the constituent vector modes [54]. Using LP modes facilitates GD characterization. Degenerate LP modes are measured to have similar GDs because the constituent vector modes are combined with equal weights and strong intragroup coupling between degenerate LP modes causes their GDs to approach the average GD of the constituent vector modes. Hence, the impulse response of an LP mode group is observed as a single narrow pulse [28], [35], [56], [57]. To date, most GD characterization experiments have studied the LP modes, using methods such as spatially and spectrally resolved imaging [56], [58]. In Section III-B, we compile experimentally measured GD spreads of LP modes in various fibers from the literature.

It is also possible to characterize the end-to-end coupled GDs $(\tau_1^{(t)}, \ldots, \tau_D^{(t)})$ of a system. The coupled GDs can be obtained as eigenvalues of the GD operator (1) computed from the end-to-end channel transfer matrix $\mathbf{M}(\Omega)$. The channel transfer matrix can be estimated by launching linear combinations of vector or LP modes and measuring the corresponding output signals [51]. Alternatively, the coupled GDs can be measured by launching the PMs, which can be computed as the eigenvectors of the GD operator or can be adaptively generated to obtain a maximally flat magnitude response [16], [51].

B. Optimized Fiber Design

The index profile of a fiber determines its uncoupled modal GDs $(\tau_1, \ldots, \tau_D)$. A primary approach for minimizing end-to-end GD spread is to design the index profile to minimize the uncoupled GD spread. For step-index fibers supporting two mode groups ($D = 6$), in theory, an arbitrarily low GD spread can be obtained by choosing a core radius at which the GD-versus-radius curves for the two mode groups intersect [59]–[61]. For fibers supporting more than two mode groups ($D > 6$), this approach does not scale easily, because the curves for different pairs of modes intersect at different radii. On the other hand, graded-index fibers with large cores ($D \to \infty$) are known to have asymptotically low GD spreads [62]. For $D \sim 12-30$ modes, similar behavior can be achieved using a depressed cladding [13], [60], [61], [64], [65].

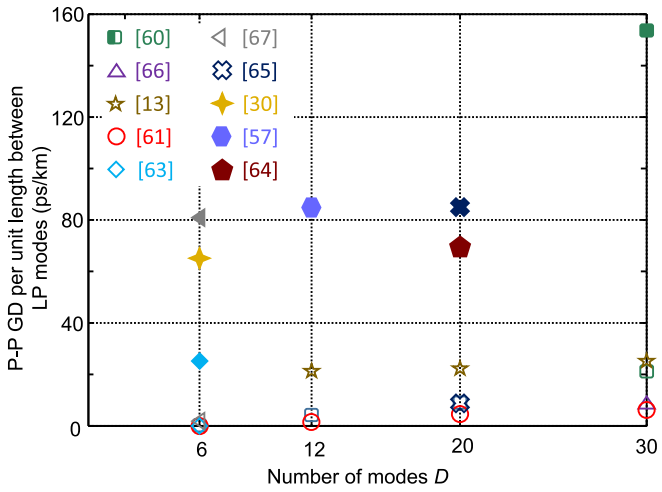

Fig. 2. Peak-to-peak GD spreads per unit length for LP modes in MMFs from the literature near 1550-nm wavelength. Open symbols show simulation results, while solid symbols show experimental results.

Fig. 2 shows *p–p* GD spread values per unit length in various MMFs presented in the literature. These have been measured for LP modes, and are defined as the maximum GD difference between any two LP mode groups, normalized by propagation length. The GD values in Fig. 2 are measured near 1550 nm. Over the entire C+L band (∼1530–1625 nm), GD spreads may be several times larger than those shown, because of mode-dependent CD. Theoretical calculations (open symbols in Fig. 2) exhibit GD spreads close to zero for $D = 6$ modes, and less than 10 ps/km up to $D = 30$ modes. However, experimental measurements (solid symbols in Fig. 3) exhibit GD spreads roughly an order of magnitude higher than the theoretical values.

The substantially higher experimental GD spreads can be attributed mainly to imperfections in fabrication [58], [66], [67]. Unlike other fiber properties, such as CD constant, effective area or attenuation, the individual uncoupled GD values are highly sensitive to index variations [58]. Hence, optimized fiber design has not proven sufficient to reduce GD spreads in fibers supporting more than two mode groups. MDM experiments for $D > 6$, e.g., [36], have encountered GD spreads that are too high to enable realization of MIMO DSP with low complexity.

C. Mode Coupling

When modes couple with each other during propagation, continuous intermixing reduces the accumulation of GD spread [68]. Designing an MDM system with strong mode coupling may be an effective strategy for minimizing the end-to-end GD spread, and may also minimize the end-to-end MDL [69], the outage probability (by enabling frequency diversity [70]), and nonlinear signal distortion [71], [72]. Strong mode coupling is currently the main technique for PMD reduction in SMF systems [73].

The effectiveness of strong mode coupling in reducing the GD spread can be viewed from multiple perspectives. As shown in Section II, the end-to-end GD spread scales with the square-root of the propagation length and the square-root of the coupling length. Hence, end-to-end GD spreads can be decreased by reducing the coupling length. The end-to-end GD spread depends only on the r.m.s. GD of each fiber section, and the impact of non-idealities in fiber fabrication can be reduced because locally high GDs can be averaged out, for example, when the GD with largest magnitude has a different sign in different sections of a fiber. Furthermore, scaling to higher D is advantageous. In the weak-coupling regime, the end-to-end GD spread depends on the uncoupled *p–p* GD. According to Fig. 2, when scaling up D, the uncoupled *p–p* GD might increase, decrease or remain approximately unchanged. In the strong-coupling regime, the end-to-end GD spread depends on the uncoupled r.m.s. GD and the ratio of the coupled *p–p* GD to the coupled r.m.s. GD. When scaling up D, although the *p–p* GD might not decrease, the r.m.s. uncoupled GD may decrease, e.g., for the GIGDC fiber [13], because computation of the r.m.s. GD includes an averaging of similar GDs. Also, the ratio of the coupled *p–p* GD to the coupled r.m.s. GD slightly decreases with D [47]. Strong mode coupling is favorable for scaling to high D.

In an MDM system, mode coupling between any two modes depends on the difference between their propagation constants, the spatial overlap between their mode profiles, and the longitudinal and transverse spatial dependence of the random or intentional perturbations inducing the coupling [6], [27], [50]. The degeneracies within mode groups play an important role in governing coupling. Most random perturbations have a slow longitudinal variation, so they strongly couple modes having nearly equal propagation constants (intragroup coupling), but weakly couple modes having highly unequal propagation constants (intergroup coupling) [6], [27].

Intragroup coupling occurs significantly over distances typically less than 1 km, e.g., after ∼1 m [74], ∼200 m [75] and ∼500 m [76]. Based on the theory in Section II, in the strong-coupling regime with a section length of 1 km, the GD spread can be reduced roughly 20-fold compared to the weak-coupling regime. Hence, intragroup coupling lengths are already close to the desired small values. The intragroup coupling strength can be increased further by intentionally introducing distributed perturbations, e.g., by spinning the MMF during the pulling process [77], [78], similar to the spinning used to reduce PMD in SMF [73]. The spin amplitude (or other parameters of intentional distributed perturbations) should be chosen carefully to minimize the GD spread [78]. For example, an excessive spin amplitude may reduce existing intragroup coupling, causing the coupled GD spread to increase [78].

Intergroup coupling, unfortunately, typically occurs over distances much longer than 1 km, and may remain small even after a full span (∼100 km) of propagation [28]. The strength of intergroup coupling depends on several variables. For example, fiber splices with mismatches of 10% to 20% of the core radius can cause considerable coupling between neighboring mode groups [79]. Coupling characteristics differ between various mode groups. Typically, higher-order mode groups have stronger intergroup coupling [80], in part, because of a larger number of possible coupling pairs. In published MDM experiments, significant intergroup coupling was demonstrated after ∼2 km [31], [80], ∼5 km [60], ∼13 km [66], ∼17 km [81], ∼40 km [29] and ∼50 km [30].

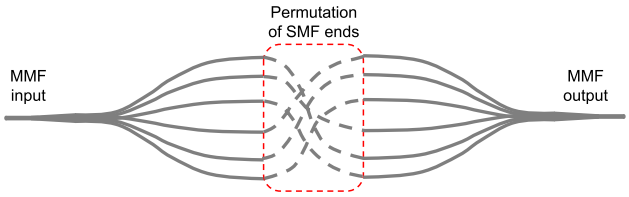

Fig. 3. Schematic of a mode scrambler using two photonic lanterns for $D = 12$ modes (six spatial modes), showing coupling from first to fifth, second to fourth, third to sixth, fourth to first, fifth to second and sixth to third spatial modes.

It may be imprudent to design an MDM system relying solely on random perturbations to reduce the GD spread (and MDL and nonlinear effects), since random perturbations are of variable effectiveness in inducing intergroup coupling. An ideal approach would be to incorporate an appropriate distributed perturbation in the fiber to increase intergroup coupling with low loss. An alternative and perhaps more practical approach is to insert lumped mode scramblers periodically along the fiber. Two classes of approaches are possible for realizing low-loss lumped mode scramblers.

The first class of approaches is based on modal spatial profiles. For example, one may interconnect a photonic lantern demultiplexer and multiplexer in series, with a permutation of the interconnecting SMFs, as depicted in Fig. 3. In a photonic lantern, modes of an MMF are adiabatically coupled to individual modes of the SMF array (also known as supermodes) by continuous evolution of the propagation constants and modal profiles. By permuted interconnection of the SMF ends corresponding to different mode groups, efficient intergroup coupling can be introduced. In theory, with adiabatic tapering, a lossless transition between guided MMF modes and separated SMFs modes is achievable. However, because of a limited photonic lantern length, insufficient index contrast with the outside substrate and manufacturing non-idealities, photonic lanterns have both insertion loss and MDL [82]. Simulations demonstrate an insertion loss of ~0.5 dB and p–p MDL of ~0.5 dB, while experiments demonstrate an insertion loss of ~4 dB and p–p MDL of 3.5 dB [82]. Loss values are doubled in a scrambler using two lanterns.

The second class of approaches is based on phase matching of modal propagation constants. For example, LPFGs are all-fiber devices having periodic perturbations, which can be written directly into photosensitized fibers [83]–[85]. The schematic of an LPFG designed for $D = 12$ modes is shown in Fig. 4(a). The device incorporates several tilted grating sections. A high tilt angle is used to obtain sufficient spatial overlap between higher-order modes. The effective grating periods are chosen to obtain phase matching between different mode groups. Phase matching between the highest-order guided modes and radiation modes is undesired, as it induces loss and MDL [84]. In each LPFG section, the grating period and length are adjusted to optimize overall coupling while minimizing losses. The LPFG is not required to produce intragroup coupling, since that is assumed provided by the transmission fiber. Fig. 4(b) shows the evolution of the power per mode in the three different mode groups as

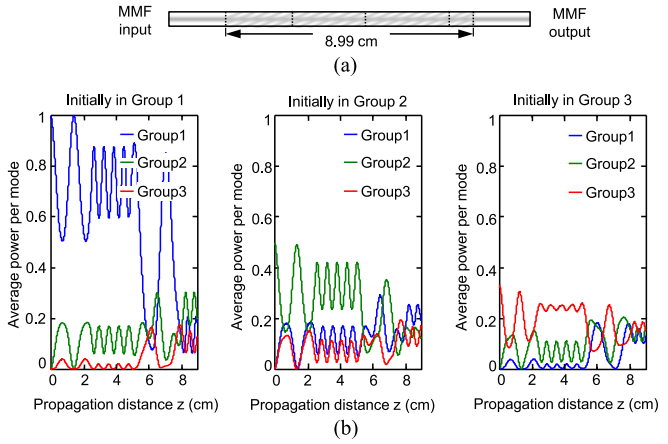

Fig. 4. (a) Schematic of a designed four-section LPFG mode scrambler for $D = 12$ modes (6 spatial modes). (b) Evolution of power per spatial mode at 1550 nm, assuming the power is initially all in one of the three mode groups (comprising one, two and three spatial modes respectively) [83].

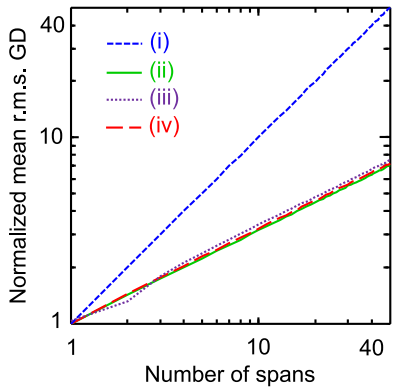

Fig. 5. The averaged normalized r.m.s. GD versus number of spans for systems with: (i) only random intragroup coupling, (ii) random intragroup coupling with random intergroup coupling (with intergroup coupling length of 100 km), (iii) random intragroup coupling with photonic lantern mode scramblers (assuming ideal photonic lanterns with zero MDL) spaced 100 km apart, and (iv) random intragroup coupling with LPFG mode scramblers (assuming the LPFGs described in Fig. 5) spaced 100 km apart. All assume an intragroup coupling length $h_{\mathrm{intra}} = 1$ km, $L_{\mathrm{amp}} = 100$ km, $D = 12$ modes and GIGDC fiber GD values. The curves are obtained using Monte-Carlo simulations of the Jones space representation for mode coupling, averaged over 300 000 random realizations.

signals propagate through the LPFG. In each four-section LPFG, simulations show an average insertion loss of 0.3 dB and r.m.s. MDL of 0.28 dB. Experimentally, LPFG mode scramblers with insertion loss less than 1.4 dB were demonstrated for $D = 6$ modes [86].

Monte-Carlo simulations are used to demonstrate the effectiveness of lumped mode scramblers in inducing strong mode coupling. Fig. 5 shows the averaged normalized r.m.s. GDs for systems with (i) only random intragroup coupling, (ii) random intragroup coupling and random intergroup coupling with 100-km coupling length, (iii) random intragroup coupling and ideal photonic lantern mode scramblers (with zero MDL) spaced 100 km apart, and (iv) random intragroup coupling with non-ideal LPFG mode scramblers (as described in Fig. 4) spaced

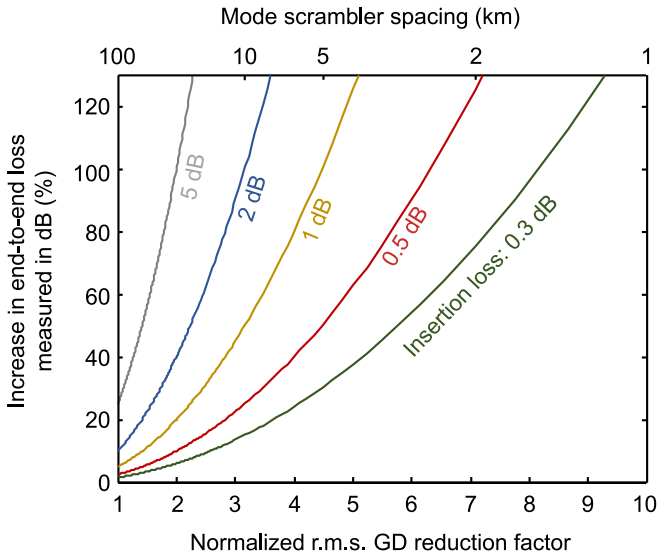

Fig. 6. Increase in end-to-end loss measured in dB (%) versus normalized r.m.s. GD reduction factor and mode scrambler spacing for various mode scrambler insertion loss values. An MMF attenuation of 0.2 dB/km is assumed.

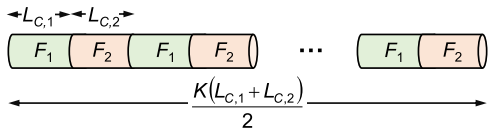

Fig. 7. Schematic of a GD-compensated system employing fiber types F_1 and F_2 with segment lengths $L_{C,1}$ and $L_{C,2}$.

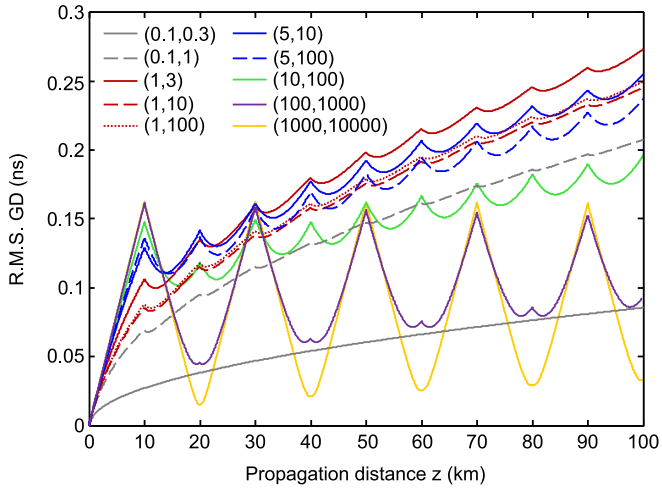

Fig. 8. Evolution of r.m.s. GD versus propagation distance z for $D = 12$ modes in a 100-km fiber span comprising ten GD compensation segments of length $L_{C,1} = L_{C,2} = 10$ km [47]. Fiber type F_1 has a GIGDC index profile and the fiber type F_2 is assumed to perfectly compensate for F_1. The legends correspond to different intra- and inter-group coupling lengths (h_{intra}, h_{inter}) in kilometers.

100 km apart. As can be observed, both (iii) and (iv) yield curves very close to (ii). Both mode scramblers are effective in inducing strong mode coupling, with an equivalent section length L_{sec} equal to the mode scrambler spacing.

In principle, by reducing the mode scrambler spacing to approach the intragroup coupling length, the GD spread can be reduced. However, limitations arise from the additional loss and cost of mode scramblers. Fig. 6 shows the increase in the end-to-end loss measured in decibel versus the normalized r.m.s. GD reduction factor. Assuming mode scramblers with 0.5-dB insertion loss and an MMF with 0.2 dB/km attenuation, a mode scrambler spacing of 100 km increases the overall loss by 2.5%, 10 km increases it by 25%, and 1 km increases it by 250%. With a goal of approaching ideal distributed mode coupling, a transmission line comprising a series of interconnected LPFGs, each randomly rotated to ensure intragroup coupling, was proposed in [85]. Such a transmission line may be expected to increase system cost and losses significantly, however.

D. GD Compensation

The end-to-end GD spread can be minimized by cascading segments of MMF in which low- and high-order modes exhibit opposing ordering of GDs. Ideally without mode coupling, the end-to-end GD spread can be reduced to zero. This can be achieved, for example as shown in Fig. 7, by concatenating segments of 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 $(\tau_{2,1}, \tau_{2,2}, \ldots, \tau_{2,D}) = -(L_{C,1}/L_{C,2})(\tau_{1,1}, \tau_{1,2}, \ldots, \tau_{1,D})$.

In experiments to date, GD compensation has enabled realization of smaller end-to-end GD spreads than using optimized single-fiber designs. Demonstrated values of the *p–p* GD spread between LP modes (see Section III-A) were 4 ps/km for $D = 6$ (using two fiber types) [87], 50 ps/km for $D = 12$ (using three fiber types) [57], and 50 ps/km for $D = 20$ (using three fiber types) [64].

There are several challenges for the implementation of GD compensation.

A first challenge is that fabricated MMFs exhibit high variations in GD values (see Section III-B). Each fiber segment must be characterized individually and matched to its compensating segments.

A second challenge is scaling to a large number of modes. For obtaining fibers with opposing GD ordering, various systematic index profile perturbations have been proposed, including modifications in the core radius, index difference, and index exponent [57], [64], [87], [88]. In practice, for more than two mode groups ($D > 6$ modes), obtaining precise GD cancellation using two, or even more, fiber types with different index profiles is likely to be challenging. When precise GD cancellation is not possible, the compensation lengths can be optimized to minimize the GD spread [48].

A third challenge for GD compensation is random mode coupling caused by perturbations and splices. For quantitative analysis of the evolution of the GD for a system in which GD compensation and mode coupling coexist, the generalized Stokes space representation is convenient [48]. The differential equations (6) and (7) are integrated assuming a piecewise constant deterministic $\partial\boldsymbol{\beta}/\partial\Omega$ corresponding to the concatenation of different fiber types. Fig. 8 shows the evolution of the r.m.s.

GD in a GD-compensated system for various coupling strengths, which are described by coupling lengths h_{intra} and h_{inter}. Fig. 8 is computed assuming two fiber types with precisely opposing GD profiles, as in Fig. 7.

When mode coupling is very weak, GD compensation is highly effective in minimizing the end-to-end GD. For the values considered in Fig. 8, this requires intergroup coupling length h_{inter} to be three to four orders of magnitude larger than the fiber segment lengths $L_{C,1}$ and $L_{C,2}$, and also requires intragroup coupling length h_{intra} to be two to three orders of magnitude larger than $L_{C,1}$ and $L_{C,2}$.

When mode coupling is strong, GD compensation becomes ineffective, and the precise ordering of GD values becomes immaterial. As explained in Section II, end-to-end GD spread depends only on the r.m.s. GD spread in each fiber segment. For the values considered in Fig. 8, 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 order to control the complexity of cable manufacturing and installation, the segment lengths $L_{C,1}$ and $L_{C,2}$ presumably cannot be much smaller than the splice spacing in current systems, which is about 1–10 km [89]. As explained in Section III-C, experiments and estimations suggest that h_{intra} should be of order 0.1–1 km, while h_{inter} should be of order 10–100 km. Given these values, Fig. 8 suggests that GD compensation is likely to be only partially effective in reducing the GD spread in deployed systems.

To further reduce the GD spread, the GD compensation approach can be combined with the mode coupling approach by inserting mode scramblers between GD-compensated fiber spans [48]. As an example, we consider an exemplary 20×100 km system that is comprised of GIGDC fibers [13]. When one LPFG mode scrambler is inserted in each 100-km span, the simulated end-to-end GD spread is $\sim$24 ns. When the LPFG mode scramblers are inserted at 10-km intervals, the end-to-end GD spread is reduced to $\sim$8.5 ns. When one LPFG mode scramblers is inserted in each 100-km span and the spans are precisely GD-compensated with segment lengths $L_{C,1} = L_{C,2} = 10$ km, the simulated end-to-end GD spread is $\sim$6 ns, assuming the worst-case random mode coupling.

IV. MIMO Signal Processing

A. Architectures and Algorithms

Receivers in SMF systems, after downconversion and high-speed analog-to-digital conversion, employ several DSP functions to recover the signal [19]–[21]. Digital equalization compensates for linear channel effects, including CD, PMD, and polarization crosstalk. Other functions implemented digitally include timing and carrier recovery, and error-correction decoding. In receivers for MDM in MMF, most of these latter functions may be adapted from SMF with straightforward modification. However, compensation of MD and modal crosstalk requires different architectures than in SMF systems because of the increased MIMO dimensionality ($D \times D$ versus 2×2) and because MD in MMF typically has a GD spread far longer than PMD in SMF.

For systems with long impulse response duration, frequency-domain equalization (FDE) in single-carrier modulation yields lower complexity than time-domain equalization (TDE) because frequency-domain multiplication has lower complexity than time-domain convolution, and time–frequency domain conversions can be efficiently implemented using the fast Fourier transform (FFT) algorithm [14].

In SMF systems, CD with a long impulse response duration is typically compensated by FDE, which is programmable but static, because the CD transfer function is nearly independent of time except its slight temperature dependence [90]. PMD, polarization crosstalk, and the residual CD, with a short impulse response duration, may be compensated by 2×2 TDE, which must be adaptive, because the PMD and polarization crosstalk transfer function varies rapidly over time. In an SMF receiver, computational complexity is minimized by separately implementing two static FDEs for CD and a single 2×2 adaptive TDE or FDE.

In MDM systems, the main difference is the long impulse response duration of MD and modal crosstalk. Their compensation requires a $D \times D$ equalizer with duration (measured in samples) of

$$N_{\mathrm{MD}} = \lceil \tau_{p-p}\, r_{\mathrm{os}} R_s \rceil\,, \tag{12}$$

where R_s is the symbol rate and r_{os} is the receiver oversampling ratio, assumed to be $r_{\mathrm{os}} = 2$ throughout this paper. A set of D static FDEs for CD and a single $D \times D$ adaptive FDE for MD and modal crosstalk (and any potential residual CD) can be employed [13], and minimizes the complexity of adaptation [14].

Adaptive FDEs can be classified into two types, depending on whether or not they employ a cyclic prefix (CP) to accommodate the time-domain circular convolution property of the FFT. Using this classification, adaptive FDE algorithms from the MDM literature are listed in Table I, along with their adaptation times and computational complexities. The adaptation time T_{adapt} is defined as the total time required to adapt to an unknown channel, assuming n_{tr} blocks of known or estimated symbols are required for training until convergence. The computational complexity is defined as the number of complex multiplications per symbol in a training block. For simplicity, all algorithms in Table I assume a radix-2 FFT, which requires $N_{\mathrm{FFT}} \log_2(N_{\mathrm{FFT}})/2$ complex multiplications per block of length N_{FFT}. For other FFT radices, the number of operations is still of the order $N_{\mathrm{FFT}} log_2(N_{\mathrm{FFT}})$ for large N_{FFT}. Adaptive FDE algorithms may possibly share the first frequency domain conversion stage with CD compensation, depending on the block lengths; however, we ignore this for simplicity in the comparison.

The first class of algorithms does not employ a CP, and uses windowed time–frequency conversions to equalize an approximated channel transfer function in the frequency domain. Various implementation options are possible for this class of algorithms.

A common FDE approach without CP is based on overlap-save implementation of frequency-domain filtering [91], [92]. The block length N_{FFT} should be at least as long as the delay spread N_{MD}. To enable $N_{\mathrm{FFT}} = N_{\mathrm{MD}}$ and simplify the

TABLE I
ADAPTIVE FDE TECHNIQUES AND THEIR COMPUTATIONAL COMPLEXITIES

	Adaptive FDE technique	FFT block length N_{FFT}	Adaptation time T_{adapt}	Number of complex multiplications per symbol in a training block
Without CP	LMS-adapted with gradient constraint [93]	N_{MD}	$2n_{\mathrm{tr}}N_{\mathrm{MD}}/R_s r_{\mathrm{os}}$	$(4+4D)log_2(N_{\mathrm{FFT}})+8D$
	LMS-adapted without gradient constraint [95]	N_{MD}	$2n_{\mathrm{tr}}N_{\mathrm{MD}}/R_s r_{\mathrm{os}}$	$4log_2(N_{\mathrm{FFT}})+8D$
	Signal PSD-directed with gradient constraint [94]	N_{MD}	$2n_{\mathrm{tr}}N_{\mathrm{MD}}/R_s r_{\mathrm{os}}$	$(4+4D)log_2(N_{\mathrm{FFT}})+12D+2$
	Noise PSD-directed with gradient constraint [94]	N_{MD}	$2n_{\mathrm{tr}}N_{\mathrm{MD}}/R_s r_{\mathrm{os}}$	$(4+4D)log_2(N_{\mathrm{FFT}})+20D+4$
	Frequency domain independent component analysis [96]	N_{MD}	$n_{\mathrm{tr}}N_{\mathrm{MD}}/R_s r_{\mathrm{os}}$	$2log_2(N_{\mathrm{FFT}})+N_{\mathrm{FFT}}+3D+5$
With CP	LMS-adapted [13]	$\frac{\eta_{\mathrm{CP}}}{1-\eta_{\mathrm{CP}}}N_{\mathrm{MD}}$	$\frac{n_{\mathrm{tr}}N_{\mathrm{MD}}}{(1-\eta_{\mathrm{CP}})R_s r_{\mathrm{os}}}$	$2log_2(N_{\mathrm{FFT}})+4D+2$
	RLS-adapted [13]	$\frac{\eta_{\mathrm{CP}}}{1-\eta_{\mathrm{CP}}}N_{\mathrm{MD}}$	$\frac{n_{\mathrm{tr}}N_{\mathrm{MD}}}{(1-\eta_{\mathrm{CP}})R_s r_{\mathrm{os}}}$	$2log_2(N_{\mathrm{FFT}})+12D+4$

architecture, an overlap rate of 0.5 is commonly used by concatenating even and odd samples of two consecutive blocks in the time domain [93], [94]. With this choice, the number of FFT operations is increased, however, the total complexity for FFT operations is reduced. Note that T_{adapt} is proportional to $2N_{\mathrm{FFT}}$, since two consecutive blocks are concatenated. For adaptive estimation of the transfer function at each frequency, the least-mean-squares (LMS) algorithm is a common choice [93]. In order to avoid distortion due to the cyclic nature of the FFT, frequency-domain gradient constraints should be enforced [92], [93]. Omitting the gradient constraints avoids time-frequency conversions to calculate the gradient error. Although the LMS algorithm without gradient constraints yields a lower adaptation complexity, it converges to a higher error rate [95]. One drawback of the standard LMS algorithm is its constant step size. The convergence speed of the LMS algorithm depends on the signal-to-noise-ratio at each frequency. By adjusting the step size based on the signal or noise power spectral density (PSD), the convergence speed can be improved, at the cost of a slightly higher computational complexity [94].

An alternative to the overlap-save approach is frequency-domain independent component analysis [96]. This technique uses only one pair of windowed FFTs for time-frequency conversions and estimates the transfer function based on a frequency-domain error, unlike overlap-save based techniques with gradient constraints. To minimize complexity, $N_{\mathrm{FFT}}=N_{\mathrm{MD}}$ is chosen and T_{adapt} is proportional to N_{FFT}, since FFT blocks are processed separately without concatenation. Adaptive estimation of the channel transfer function is based on component separation with independent component analysis. At each frequency, the separation procedure includes gradient updates for a cost function that increases super-linearly with its argument. Using this technique, the performance is improved as compared to LMS-adapted FDE with gradient constraints. However, the computational complexity is very high, especially for systems with long GD spread requiring large N_{FFT}.

The second class of algorithms relies on a CP to handle the circular nature of FFT-based convolution. Each block of length N_{FFT} is prepended with a CP of length N_{CP} before transmission. The CP length N_{CP} must be no shorter than the channel delay spread and, to minimize overhead, $N_{\mathrm{CP}}=N_{\mathrm{MD}}$ should be chosen. When using a CP, linear filtering by the channel corresponds to a multiplicative relationship between the FFTs of input and output signal blocks, which simplifies implementation of adaptive FDE. Using a CP decreases complexity, as shown in Table I, and the reduction becomes more significant for a large number of modes *D*. On the other hand, it reduces system throughput and average-power efficiency by a factor corresponding to a CP efficiency parameter:

$$\eta_{\mathrm{CP}}=\frac{N_{\mathrm{FFT}}}{N_{\mathrm{FFT}}+N_{\mathrm{MD}}}. \tag{13}$$

When using a CP, the adaptation time T_{adapt} is proportional to $N_{\mathrm{FFT}}+N_{\mathrm{MD}}$ instead of N_{FFT}, so it is increased by a factor $1/(1-\eta_{\mathrm{CP}})$. For adaptive estimation of the channel transfer function at each frequency, two options are the LMS and recursive least squares (RLS) [13], [91] algorithms. As described previously, the constant step size of the standard LMS algorithm causes slow convergence and a high error rate. To overcome this problem, RLS also tracks an estimate of an inverse autocorrelation function of the received signal to optimize the adaptation step size [13], [91]. RLS yields faster convergence and a lower error rate, at the cost of a higher computational complexity for each iteration.

The main benefit of using a CP is reduced computational complexity in a training block, and the main drawbacks are decreased throughput and average-power efficiencies, and potentially increased adaptation time T_{adapt}. On the other hand, both the total equalization complexity and total adaptation time are proportional to the required number of training blocks n_{tr}. Given a target performance metric, the required n_{tr} is impacted by the signal-to-noise ratio, MDL, the number of modes *D*, and mode coupling regime [13], [94]. A comparison of the required n_{tr} for all the algorithms in Table I has not been made to date and is an important future research topic.

B. Impact of GD Spread on Computational Complexity

The large GD spread in MDM systems necessitates implementation of $D\times D$ adaptive MIMO equalization in the frequency-domain. In adaptive FDE, the *p–p* GD spread determines the FFT block length N_{FFT}, which affects adaptation time, performance, and complexity.

In the first category of algorithms not using CP, both complexity and adaptation time increase with N_{FFT}. With the choice $N_{\mathrm{FFT}}=N_{\mathrm{MD}}$, the computational complexity per symbol to scales at a rate between $\log(N_{\mathrm{MD}})$ and N_{MD}, and the adapta-

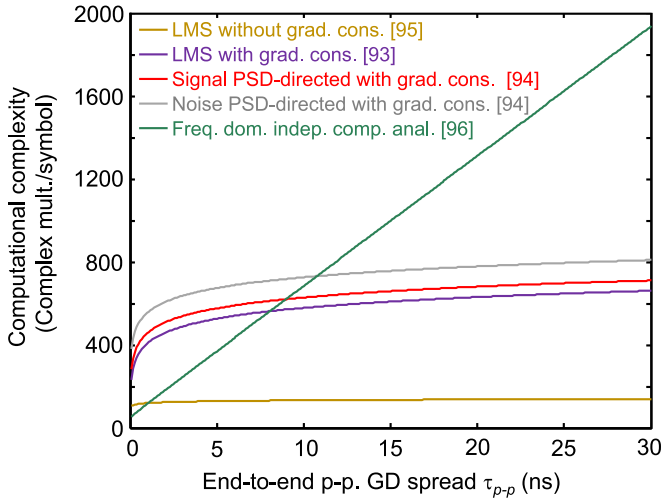

Fig. 9. Computational complexity measured in complex multiplications per symbol versus end-to-end *p–p* GD spread τ_{p-p} for adaptive FDE algorithms without CP. $D = 12$ modes, symbol rate $R_s = 32$ Gbd and oversampling ratio $r_{\rm os} = 2$ are assumed.

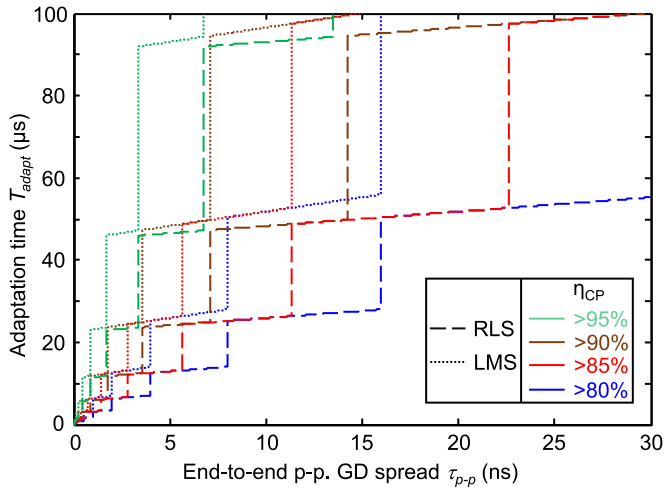

Fig. 10. Adaptation time $T_{\rm adapt}$ versus end-to-end *p–p* GD spread τ_{p-p} for LMS- (dotted lines) and RLS-adapted (dashed lines) MIMO FDE using CP. The colors denote minimum CP efficiencies and corresponding choices of FFT block length $N_{\rm FFT}$ $D = 12$ modes, symbol rate $R_s = 32$ Gbd and oversampling ratio $r_{\rm os} = 2$ are assumed. The LMS algorithm assumes $n_{\rm tr} = 700$ and the RLS algorithm assumes $n_{\rm tr} = 350$ [13].

tion time scales with $N_{\rm MD}$ (see Table I). Fig. 9 demonstrates the computational complexity versus end-to-end *p–p* GD spread for the algorithms given in Table I for $D = 12$ modes. At present, it is not possible to compare the adaptation time $T_{\rm adapt}$ required for all these algorithms, since the required number of training sequences $n_{\rm tr}$ is not known for all the algorithms.

In the second category of algorithms using CP, the choice of $N_{\rm FFT}$ depends on the trade-offs between CP efficiency, adaptation time, and computational complexity. To achieve sufficiently high CP efficiency $\eta_{\rm CP}$, the block length $N_{\rm FFT}$ should be significantly longer than the delay spread, or $N_{\rm FFT} \gg N_{\rm MD}$. On the other hand, the adaptation time $T_{\rm adapt}$ is proportional to $N_{\rm FFT}$, so fast adaptation favors a small $N_{\rm FFT}$. The computational complexity scales with $\log(N_{\rm FFT})$, hence the effect of $N_{\rm FFT}$ on complexity is less significant than its effect on CP efficiency and adaptation time. Figs. 10 and 11 show the adaptation time $T_{\rm adapt}$ and computational complexity as a function of end-to-end *p–p* GD spread, for various values of the CP efficiency $\eta_{\rm CP}$ (using $n_{\rm tr}$ values from [13]) for $D = 12$ modes. An MDM system should be designed so that $T_{\rm adapt}$ is shorter than the time scale of MMF channel variations, which might be as short as a few tens of microseconds [32], [33]. Hence, a low end-to-end GD spread is crucial for minimizing complexity while achieving sufficiently short adaptation time. Systems having high end-to-end GD spread may use RLS instead of LMS adaptation to reduce the required number of training blocks $n_{\rm tr}$, albeit at the cost of an approximate doubling of computational complexity. Systems having high end-to-end GD may also use a lower $\eta_{\rm CP}$ to reduce $N_{\rm FFT}$, albeit at the cost of a decreased power efficiency.

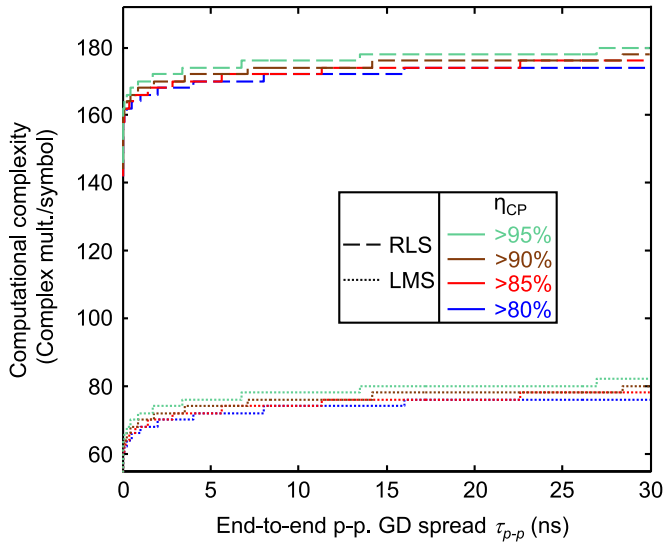

Fig. 11. Computational complexity measured in complex multiplications per symbol versus end-to-end *p–p* GD spread τ_{p-p} for LMS- (dotted lines) and RLS-adapted (dashed lines) MIMO FDE using CP. The colors denote minimum CP efficiencies and corresponding choices of FFT block length $N_{\rm FFT}$. $D = 12$ modes, symbol rate $R_s = 32$ Gbd and oversampling ratio $r_{\rm os} = 2$ are assumed.

V. Discussion and Future Work

We have summarized known approaches for managing end-to-end GD spread and optimizing the performance and complexity of MIMO DSP. Significant work is needed in both areas to make MDM a practical reality. Here, we discuss important directions for future work.

The GD spreads obtained to date solely through optimized fiber design are not sufficiently low, especially for $D > 6$ modes. Fabricated fibers have far higher GD spreads than predicted by theory, presumably owing to imperfections in fabrication processes. The significant variations within and between different sections of fabricated fibers are problematic for the large-scale deployment of MDM. It appears likely that GD management must rely substantially on mode coupling and/or GD compensation.

Mode coupling is a promising approach for reduction of end-to-end GD spread. Periodic insertion of mode scramblers based on photonic lanterns or LPFGs, in conjunction with random intragroup coupling, is sufficient to induce strong mode coupling, which can reduce GD spread, as well as MDL and nonlinear effects. Remaining concerns for mode scrambler devices are excessive insertion loss, MDL, and cost. The achievable loss, MDL and cost will determine at what intervals the devices can be inserted, and what reductions of GD, MDL and nonlinearity can be obtained. As an alternative to lumped mode scramblers, it would be desirable to introduce distributed perturbations during MMF manufacturing to induce strong spatial mode coupling, similar to the "spinning" used to reduce PMD in SMF. If such perturbations can achieve strong coupling with low loss, they would be preferable to lumped mode scramblers.

GD compensation is only partially effective in minimizing GD spread in the presence of mode coupling caused by random perturbations and splices. To increase the effectiveness of GD compensation, short GD compensation sections should be used, but this may increase system cost. Also, GD compensation is only partially effective when the different fiber types do not compensate each other perfectly, whether due to limitations in design or non-ideality in fabrication. Continued improvements in fiber design, fabrication and characterization can help advance GD compensation.

MIMO DSP complexity per symbol in MDM is higher than that in SMF systems, even with optimized architectures. It is desirable to use an adaptive FDE algorithm yielding the minimum complexity given performance and adaptation time requirements. For the algorithms listed in Table I, higher-complexity algorithms typically yield faster convergence and lower converged error ratio. Due to uncertainties in eventual deployment scenarios, which may affect performance requirements, the optimal choice of algorithm for commercial systems is difficult to determine at present. The adaptive FDE algorithms are designed to achieve high performance in dynamic channel environments. Although experimental studies suggest time scales of tens of microseconds for MDM channel dynamics, a deeper understanding of the underlying physics and development of accurate dynamic channel models are important for testing of MIMO DSP algorithms. Better knowledge of dynamic channel characteristics may enable further optimization of the algorithms.

VI. Conclusion

We have reviewed and compared generalized Jones and Stokes representations for GD modeling. We have discussed three GD management techniques: optimized fiber design, mode coupling, and GD compensation. MDM systems cannot rely merely on optimized fiber design for GD spread minimization, because non-ideal fabrication significantly increases the GD spread. Strong mode coupling can be very effective in reducing the GD spread. Although random perturbations in MDM systems cause sufficient intragroup coupling, intergroup coupling is typically weak and may need to be enhanced by intentionally introducing distributed perturbations or periodic insertion of lumped mode scramblers. Two mode scrambler designs, based on photonic lanterns and LPFGs, are both very effective in inducing strong mode coupling. GD compensation is only partially effective in reducing the GD spread, due to difficulties in designing and fabricating precisely compensating fibers and due to random perturbations that induce mode coupling.

A long GD spread in MDM systems necessitates adaptive MIMO equalization in the frequency domain. For adaptive FDEs with or without CP, we described various adaptive FDE algorithms and compared their trade-offs between complexity, convergence speed, and performance. Higher-complexity algorithms typically achieve faster convergence and lower converged error ratio. While having a lower complexity, FDE with CP yields a lower throughput efficiency and a somewhat slower convergence. Convergence speed can be increased by decreasing the CP efficiency, which reduces throughput and average-power efficiency. Because different algorithms exhibit various advantages, further work is needed to identify and validate the optimal choice for eventual deployment. For all known adaptive FDE algorithms, managing the system's end-to-end GD spread is crucial to control the complexity, adaptation time, and performance of MIMO DSP.

References

[1] R.-J. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks," *Proc. IEEE*, vol. 100, no. 5, pp. 1035–1055, May 2012.

[2] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik, "On achievable rates for long-haul fiber-optic communications," *Opt. Exp.*, vol. 23, no. 7, pp. 9183–9191, Apr. 2015.

[3] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.

[4] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[5] 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.

[6] 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.

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

[8] D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," *Nature Photon.*, vol. 7, no. 5, pp. 354–362, Apr. 2013.

[9] G. Li, N. Bai, N. Zhao, and C. Xia, "Space-division multiplexing: The next frontier in optical communication," *Adv. Opt. Photon.*, vol. 6, no. 4, pp. 413–487, Dec. 2014.

[10] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," *Opt. Exp.*, vol. 19, no. 17, pp. 16680–11696, Aug. 2011.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] A. Mecozzi, C. Antonelli, and M. Shtaif, "Intensity impulse response of SDM links," *Opt. Exp.*, vol. 23, no. 5, pp. 5738–5743, Feb. 2015.

[16] J. Carpenter, B. J. Eggleton, and J. Schröder, "Observation of Eisenbud–Wigner–Smith states as principal modes in multimode fibre," *Nature Photon.*, vol. 9, no. 11, pp. 751–757, Oct. 2015.

[17] R. Ryf *et al.*, "Space-division multiplexed transmission over 4,200 km 3-core microstructured fiber," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2012, paper PDP5C.2.

[18] T. Fujisawa and K. Saitoh, "Impulse response analysis of strongly-coupled three-core fibers," in *Proc. Front. Opt.*, San Jose, CA, USA, 2015, paper FM1E-3.

[19] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1164–1179, Oct. 2010.

[20] 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.

[21] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," *IEEE J. Quantum Electron.*, vol. 16, no. 5, pp. 1180–1192, Oct. 2010.

[22] P. M. Krummrich and K. Kotten, "Extremely fast (microsecond scale) polarization changes in high speed long haul WDM transmission systems," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2004, paper FI3.

[23] T. Pfau *et al.*, "Ultra-fast adaptive digital polarization control in a real-time coherent polarization-multiplexed QPSK receiver," presented at the Optical Fiber Communication Conf., San Diego, CA, USA, 2008, paper OTuM3.

[24] B. Pillai *et al.*, "End-to-end energy modeling and analysis of long-haul coherent transmission systems," *J. Lightw. Technol.*, vol. 32, no. 18, pp. 3093–3111, Sep. 2014.

[25] B. Inan *et al.*, "DSP complexity of mode-division multiplexed receivers," *Opt. Exp.*, vol. 20, no. 10, pp. 10859–10869, May 2012.

[26] T. Sakamoto *et al.*, "Mode-division multiplexing transmission system with DMD-independent low complexity MIMO processing," *J. Lightw. Technol.*, vol. 31, no. 13, pp. 2192–2199, Jul. 2013.

[27] 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.

[28] 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.

[29] P. Genevaux *et al.*, "Comparison of QPSK and 8-QAM in a three spatial modes transmission," *IEEE Photon. Technol. Lett.*, vol. 26, no. 4, pp. 414–417, Feb. 2014.

[30] N. Bai *et al.*, "Mode-division multiplexed transmission with inline few-mode fiber amplifier," *Opt. Exp.*, vol. 20, no. 3, pp. 2668–2680, Jan. 2012.

[31] S. H. Chang *et al.*, "Mode division multiplexed optical transmission enabled by all-fiber mode multiplexer," *Opt. Exp.*, vol. 22, no. 12, pp. 14229–14236, Jun. 2014.

[32] X. Chen, J. He, A. Li, J. Ye, and W. Shieh, "Characterization of dynamic evolution of channel matrix in two-mode fibers," presented at the Optical Fiber Communication Conf., Anaheim, CA, USA, 2013, paper OM2C-3.

[33] X. Chen, J. He, A. Li, J. Ye, and W. Shieh, "Characterization and analysis of few-mode fiber channel dynamics," *IEEE Photon. Technol. Lett.*, vol. 25, no. 18, pp. 1819–1822, Sep. 2013.

[34] R. Ryf *et al.*, "12 × 12 MIMO transmission over 130-km few-mode fiber," presented at the Frontiers Optics Conf., Rochester, NY, USA, 2012, paper FW6C.4.

[35] V. A. J. M. Sleiffer *et al.*, "73.7 Tb/s (96 x 3 x 256-Gb/s) mode-division-multiplexed DP-16QAM transmission with inline MM-EDFA," *Opt. Exp.*, vol. 20, no. 26, pp. B428–B438, Dec. 2012.

[36] N. K. Fontaine *et al.*, "30×30 MIMO transmission over 15 spatial modes," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2015, paper Th5C.1.

[37] R. van Uden, C. Okonkwo, H. Chen, H. de Waardt, and A. Koonen, "6×28GBaud 128-SP-QAM Transmission over 41.7 km Few-Mode Fiber with a 6×6 MIMO FDE," presented at the Optical Fiber Communication Conf., San Francisco, CA, USA, 2014, paper W4J.4.

[38] S. Randel *et al.*, "First real-time coherent MIMO-DSP for six coupled mode transmission," in *Proc. Photon. Conf.*, 2015, pp. 1–2.

[39] T. Hayashi *et al.*, "Coupled-core multi-core fibers: High-spatial-density optical transmission fibers with low differential modal properties," in *Proc. Eur. Conf. Opt. Commun.*, 2015, pp. 1–3.

[40] S. O. Arik and J. M. Kahn, "Coupled-core multi-core fiber for spatial multiplexing," *IEEE Photon. Technol. Lett.*, vol. 25, no. 21, pp. 2054–2057, Nov. 2013.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] S. Fan and J. M. Kahn, "Principal modes in multi-mode waveguides," *Opt. Lett.*, vol. 30, no. 2, pp. 135–137, Jan. 2005.

[47] K.-P. Ho and J. M. Kahn, "Delay-spread distribution for multimode fiber with strong mode coupling," *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1906–1909, Nov. 2012.

[48] S. O. Arik, K. P. Ho, and J. M. Kahn, "Delay spread reduction in mode-division multiplexing: Mode coupling vs. delay compensation," *J. Lightw. Technol.*, vol. 33, no. 21, pp. 4504–4512, Nov. 2015.

[49] 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.

[50] 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.

[51] J. Carpenter, B. J. Eggleton, and J. Schröder, "First demonstration of principal modes in multimode fiber," presented at the Eur. Conf. Optical Communication, Cannes, France, 2014, pp. PD.2.1.

[52] R. Brüning, D. Flamm, L. Lukas, J. Lenz, and M. Duparré, "Determination of the physical fiber modes," *Proc. SPIE Next-Gener. Opt. Commun., Compon., Sub-Syst., Syst. IV*, vol. 9389, pp. 93890O1–93890O9, Feb. 2015.

[53] B. Ung *et al.*, "Inverse-parabolic graded-index profile for transmission of cylindrical vector modes in optical fibers," presented at the Optical Fiber Communication Conf., San Francisco, CA, USA, 2014, paper Tu3K.4.

[54] H. Kogelnik and P. J. Winzer, "Modal birefringence in weakly guiding fibers," *J. Lightw. Technol.*, vol. 30, no. 14, pp. 2240–2245, Jul. 2012.

[55] A. W. Snyder and W. R. Young, "Modes of optical waveguides," *J. Opt. Soc. Am.*, vol. 68, pp. 297–309, Mar. 1978.

[56] J. W. Nicholson, A. D. Yablon, S. Ramachandran, and S. Ghalmi, "Spatially and spectrally resolved imaging of modal content in large-mode-area fibers," *Opt. Exp.*, vol. 16, no. 10, pp. 7233–7243, May 2008.

[57] T. Mori, T. Sakamoto, M. Wada, T. Yamamoto, and F. Yamamoto, "Low DMD four LP mode transmission fiber for wide-band WDM-MIMO system," presented at the Optical Fiber Communication Conf., Anaheim, CA, USA, 2013, paper OTh3K-1.

[58] L. G. Nielsen, Y. Sun, R. Vincentz, S. Jensen, J. W. Nicholson, and R. Lingle, "Recent advances in low DGD few-mode fibre design, fabrication, characterization and experiments," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2015, paper M2C-3.

[59] 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.

[60] P. Sillard, D. Molin, M. Bigot-Astruc, K. de Jongh, and F. Achten, "Low-differential-mode-group-delay 9-LP-mode fiber," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2015, paper M2C.2.

[61] F. M. Ferreira, D. Fonseca, and H. J. A. da Silva, "Design of few-mode fibers with M-modes and low differential mode delay," *J. Lightw. Technol.*, vol. 32, no. 3, pp. 353–360, Feb. 2014.

[62] 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.

[63] K. Sato, R. Maruyama, N. Kuwaki, S. Matsuo, and M. Ohashi, "Optimized graded index two-mode optical fiber with low DMD, large A_{eff} and low bending loss," *Opt. Exp.*, vol. 21, no. 14, pp. 16231–16238, Jul. 2013.

[64] T. Mori, T. Sakamoto, M. Wada, T. Yamamoto, and F. Yamamoto, "Six-LP-mode transmission fiber with DMD of less than 70 ps/km over C+L band," presented at the Optical Fiber Communication Conf., San Francisco, CA, USA, 2014, paper M3F-3.

[65] P. Sillard, M. Bigot-Astruc, and D. Molin, "Few-mode fibers for mode division-multiplexed systems," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2824–2829, Aug. 2014.
[66] R. Jensen, L. G. Nielsen, N. H. Wong, Y. Sun, Y. Jung, and D. J. Richardson, "Demonstration of a 9 LP-mode transmission fiber with low DMD and loss," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2015, paper W2A.34.
[67] L. Gruner-Nielsen *et al.*, "Few mode transmission fiber with low DGD, low mode coupling, and low loss," *J. Lightw. Technol.*, vol. 30, no. 23, pp. 3693–3698, Dec. 2012.
[68] S. D. Personick, "Time dispersion in dielectric waveguides," *Bell Sys. Tech. J.*, vol. 50, no. 3, pp. 843–859, Mar. 1971.
[69] 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.
[70] K.-P. Ho and J. M. Kahn, "Frequency diversity in mode-division multiplexing systems," *J. Lightw. Technol.*, vol. 29, no. 24, pp. 3719–3726, Dec. 2011.
[71] 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.
[72] 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.
[73] 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.
[74] R. Gabet *et al.*, "Complete dispersion characterization of few mode fibers by OLCI technique," *J. Lightw. Technol*, vol. 33, no. 6, pp. 1155–1160, Mar. 2015.
[75] S. Murshid, B. Grossman, and P. Narakorn, "Spatial domain multiplexing: A new dimension in fiber optic multiplexing," *Opt. Laser Technol.*, vol. 40, no. 8, pp. 1030–1036, Nov. 2008.
[76] L. Grüner-Nielsen *et al.*, "Measuring distributed mode scattering in few mode fibers with high and low differential group delay," in *Proc. IEEE Photon. Soc. Summer Top. Meet.*, Jul. 2012, paper TuC1.3, pp. 193–194.
[77] 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.
[78] L. Palmieri and A. Galtorossa, "Intra-modal dispersion properties of step-index few-mode spun fibers," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 303–313, Jan. 2016.
[79] J. Vuong *et al.*, "Mode coupling at connectors in mode-division multiplexed transmission over few-mode fiber," *Opt. Exp.*, vol. 23, no. 2, pp. 1438–1455, Jan. 2015.
[80] J. Carpenter, B. C. Thomsen, and T. D. Wilkinson, "Degenerate mode-group division multiplexing," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3946–3952, Dec. 15, 2012.
[81] R. Ryf *et al.*, "Mode-multiplexed transmission over conventional graded index multimode fibers," *Opt. Exp.*, vol. 23, no. 1, pp. 235–246, Jan. 2015.
[82] H. Chen *et al.*, "Design constraints of photonic-lantern spatial multiplexer based on laser-inscribed 3-D waveguide technology," *J. Lightw. Technol.*, vol. 33, no. 6, pp. 1147–1154, Mar. 2015.
[83] 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.
[84] D. Askarov and J. M. Kahn, "Long-period fiber gratings for mode coupling in mode-division-multiplexing systems," *J. Lightw. Technol.*, vol. 33, no. 19, pp. 4032–4038, Oct. 2015.
[85] J. Fang, A. Li, and W. Shieh, "Low-DMD few-mode fiber with distributed long-period grating," *Opt. Lett.*, vol. 40, no. 17, pp. 3937–3940, Sep. 2015.
[86] T. Mori *et al.*, "Equipartition multiplexing technique for equalizing channel dependent degradation in MDM transmission," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2015, paper Tu2D.2.
[87] R. Maruyama, N. Kuwaki, S. Matsuo, and M. Ohashi, "Few mode fibers with low DMD slope realizing zero-DMD in wide wavelength range for MIMO processing," *Proc. SPIE Next-Gener. Opt. Commun., Compon., Sub-Syst. Syst. III*, vol. 9009, no. 90090B, pp. 90090B1–90090B8, Dec. 2013.
[88] 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.
[89] R. J. Feuerstein, "Field measurements of deployed fibers," presented at the Nat. Fiber Optic Engineers Conf., Anaheim, CA, USA, 2005, paper. NThC4.
[90] T. Kato, Y. Koyano, and M. Nishimura, "Temperature dependence of chromatic dispersion in various types of optical fiber," *Opt. Lett.*, vol. 25, no. 16, pp. 1156–1158, Aug. 2000.
[91] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2001, no. 7, pp. 344–356.
[92] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.
[93] N. Bai and G. Li, "Adaptive frequency-domain equalization for mode division multiplexed transmission," *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1918–1921, Nov. 2012.
[94] X. He, Y. Weng, and Z. Pan, "A step-size controlled method for fast convergent adaptive FD-LMS algorithm in few-mode fiber communication systems," *J. Lightw. Technol.*, vol. 32, no. 22, pp. 3820–3826, Nov. 2014.
[95] S. Randel, P. J. Winzer, M. Montoliu, and R. Ryf, "Complexity analysis of adaptive frequency-domain equalization for MIMO-SDM transmission," in *Proc. 39th Eur. Conf. Exhib. Opt. Commun.*, Sep. 2013, pp. 1–3.
[96] L. Zhao, G. Hu, L. Yan, H. Wang, and L. Li, "Mode demultiplexing based on frequency-domain-independent component analysis," *IEEE Photon. Technol. Lett.*, vol. 27, no. 2, pp. 185–188, Jan. 2015.

Sercan Ö. Arık received the B.S. degree from Bilkent University, Ankara, Turkey, in 2011, and the M.S. degree from Stanford University, Stanford, CA, USA, in electrical engineering, in 2013, where he is currently working toward the Ph.D. degree. 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, Berkeley, CA, USA, in 1993 and 1995, respectively, all in electrical engineering. He is currently at SiBEAM, Sunnyvale, CA, (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 wireless HD 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 the 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. 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). Over the years, his major fields include high-speed digital communication systems via optical fiber, copper wire, or radio frequency.

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 (U.C. Berkeley), Berkeley, CA, USA, in 1981, 1983, and 1986, respectively. From 1987 to 1990, he was with ATamp;&T Bell Laboratories, Crawford Hill Laboratory, Holmdel, NJ, USA. He demonstrated multi-Gb/s 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 at U.C. Berkeley, performing research on optical and wireless communications. Since 2003, he has been a Professor of electrical engineering at Stanford University, Stanford, CA, USA, 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 found StrataLight Communications, where he served as the Chief Scientist from 2000 to 2003. StrataLight was acquired by Opnext, Inc., in 2009.