Linear Propagation Effects in Mode-Division Multiplexing Systems

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

(Invited Tutorial)

***Abstract*—In this paper, we review linear propagation effects in a multimode fiber (MMF) and their impact on performance and complexity in long-haul mode-division multiplexing (MDM) systems. We highlight the many similarities to wireless multi-input multi-output (MIMO) systems. Mode-dependent loss and gain (MDL), analogous to multipath fading, can reduce average channel capacity and cause outage in narrowband systems. Modal dispersion (MD), analogous to multipath delay spread, affects the complexity of MIMO equalization, but has no fundamental effect on performance. Optimal MIMO transmission uses a basis of the Schmidt modes, which may be obtained by a singular value decomposition of the MIMO channel. In the special case of a unitary channel (no MDL), an optimal basis is the set of principal modes, which are eigenvectors of a group delay operator, and are free of signal distortion to first order. We present a concatenation rule for the accumulation of MD along a multisection link. We review mode coupling in MMF, including physical origins, models, and regimes of weak and strong coupling. Strong mode coupling is a key to overcoming challenges in MDM systems. Strong coupling reduces the group delay spread from MD, minimizing the complexity of MIMO signal processing. Likewise, it reduces the variations of loss and gain from MDL, maximizing channel capacity. In the strong-coupling regime, the statistics of MD and MDL depend only on the number of modes and the variance of accumulated group delay or loss/gain, and can be derived from the eigenvalue distributions of certain Gaussian random matrices.**

***Index Terms*—Channel capacity, frequency diversity, MIMO, mode-division multiplexing, multimode fiber.**

I. Introduction

OPTICAL fiber communication systems have undergone sustained and dramatic improvements in bit rate, reach, and functionality since their origins in the 1960s [1]. In the first few decades, the bit rate per fiber was increased by exploiting physical technologies. These include low-loss fibers, active components operating at the low-loss window, such as lasers, modulators and Erbium-doped fiber amplifiers, and the use of wavelength-division multiplexing [2], [3]. During the past decade, coherent detection has been combined with advanced signal processing techniques and powerful error-correction codes to further improve transmission system performance [4], [5].

All available degrees of freedom—time, frequency, phase, and polarization—are exploited in current long-haul single-mode fiber (SMF) systems, which achieve bit rates above 100 Gb/s per channel and spectral efficiencies exceeding 2 b/s/Hz [4], [5]. These systems are fundamentally limited only by optical amplifier noise and fiber nonlinearities [6]–[8]. Technical options to further increase the capacity of SMF systems are limited, and are reaching a point of diminishing return. Future innovations in SMF systems are likely to emphasize cost reduction, ease of deployment, and management flexibility.

Mode-division multiplexing (MDM) systems use a plurality of modes in a multimode fiber (MMF) to transmit many independent parallel data streams [9]–[13]. The electric fields in an MMF can be expressed in terms of a set of orthogonal spatial and polarization modes. When all the degrees of freedom exploited in SMF are used in each spatial mode of an MDM system, the overall bit rate increases in proportion to the number of spatial modes. While long-distance propagation in MMF inevitably causes coupling between different modes, leading to crosstalk and interference [14], [15], multi-input multi-output (MIMO) signal processing techniques can separate the parallel data streams. As in MIMO wireless systems [16]–[18], the coupled parallel channels can be used to either increase the data rate or enhance system reliability, providing multiplexing or diversity gain [17], [19].

MDM and wireless systems both use electromagnetic waves. MIMO channels in these systems share many similarities, but exhibit significant differences. This paper focuses on propagation effects in MDM systems in the linear regime, where the two major effects are *modal dispersion* (MD) [20] and *mode-dependent gains and losses* (collectively referred to as MDL) [21]. These effects are also inherent in wireless MIMO systems, where they are typically described using different terminology.

MD in MDM systems is analogous to the multipath delay spread in wireless MIMO systems. In both types of systems, MIMO equalizers must have sufficient memory to compensate for the delay spread. Likewise, MDL is analogous to the multipath fading in wireless MIMO systems. In both types of systems, these effects cause variations among the gains of different spatial channels and may degrade system performance. Wireless

Manuscript received June 22, 2013; revised August 28, 2013 and September 18, 2013; accepted September 19, 2013. Date of publication September 30, 2013; date of current version January 10, 2014. The work of J. M. Kahn was supported in part by the National Science Foundation under Grant ECCS-1101905 and in part by Corning, Inc.

K.-P. Ho is with Silicon Image, Sunnyvale, CA 94085 USA (e-mail: kpho@ieee.org).

J. M. Kahn is with the E. L. Ginzton Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: jmk@ee.stanford.edu).

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.2013.2283797

MIMO signals [22], [23] often occupy a narrow frequency bandwidth, so that fading may be nearly flat (frequency-independent) for all spatial channels. By contrast, MDM signals typically occupy a very wide bandwidth, so modal gains and losses become frequency-selective, yielding efficient frequency diversity that potentially improves performance [24].

Mode coupling is typically beneficial in MDM systems [13]. Strong mode coupling reduces the delay spread from MD, minimizing MIMO signal processing complexity, and it reduces the spread of modal gains/losses from MDL, minimizing any reduction of channel capacity. Strong mode coupling also enables frequency diversity, which minimizes the probability of system outage caused by MDL. Although outside the scope of this paper, strong mode coupling also can reduce the impact of fiber nonlinearities [25], [26]. MDM systems for long-haul transmission may not achieve practical, reliable operation without strong mode coupling [13], [27].

In the strong-mode-coupling regime, end-to-end linear propagation in an MDM system may be modeled using the product of many random matrices [13], [21], [24], [28], [29]. This random matrix model is helpful in deriving the statistics of MD and MDL in the strong mode-coupling regime. Modal group delays have the same distribution as the eigenvalues of a zero-trace, zero-mean Gaussian random matrix [13], [20]. Modal gains, measured in logarithmic units, have the same distribution as the eigenvalues of a zero-trace Rician matrix [29], or a zero-trace, nonzero-mean Gaussian random matrix.

The remainder of this paper is as follows. Section II describes some general end-to-end characteristics of linear MIMO systems, which may include MDM or wireless MIMO systems, and defines MD and MDL. Section III discusses the physical origins of mode coupling in MMFs, a multisection model for propagation in MMFs, and regimes of weak and strong mode coupling. Section IV describes the statistics of MD in the strong-coupling regime and computes the delay spread, which affects signal processing complexity. Section V discusses the statistics of strongly coupled MDL and how MDL affects the channel capacity of MDM systems. Section VI presents our conclusions.

II. Linear MIMO Systems

In this section, certain fundamental end-to-end properties of linear MIMO systems are reviewed. These properties apply equally to MDM or wireless MIMO systems, although the terminology employed is somewhat specialized to MDM systems. Two important sets of eigenmodes are described. The *Schmidt modes* are eigenvectors of a *phase-conjugate round-trip propagation operator*, and may obtained directly from the singular value decomposition (SVD) of a MIMO channel. They generally represent an optimal basis for MDM transmission when using coherent detection. The *principal modes* are eigenvectors of a *group delay operator*, and may provide a suitable basis for MDM transmission when using direct detection, assuming MDL and higher-order MD are tolerably small.

A. Channel Decomposition and Optimal Transmission

We assume an MMF supporting D propagating modes, including spatial and polarization degrees of freedom.

To fully utilize the available degrees of freedom and avoid outage, an MDM system should excite and detect all D modes [13], [28]. In this case, neglecting optical amplifier noise and fiber nonlinearities [25], [26], end-to-end propagation can be described by a frequency-dependent $D \times D$ matrix

$$\mathbf{M}^{(t)}(\omega). \tag{1}$$

The matrix $\mathbf{M}^{(t)}(\omega)$ can include the effects of modal (de)multiplexers, fibers, amplifiers, reconfigurable add-drop (de)multiplexers, and other linear components in the link, and can account for arbitrary gain or loss, mode coupling, or dispersion. The input–output relationship of any noiseless, linear MIMO system can be described by a frequency-dependent matrix [13], [16], [22], [23], [28]. If unequal numbers of modes are excited and detected, $\mathbf{M}^{(t)}(\omega)$ is not a square matrix.

Using an SVD [30, Sec. 2.5] at each angular frequency ω, the propagation matrix (1) may be expressed as

$$\mathbf{M}^{(t)}(\omega) = \mathbf{V}^{(t)}(\omega)\mathbf{\Lambda}^{(t)}(\omega)\mathbf{U}^{(t)^*}(\omega) \tag{2}$$

where $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)}(\omega)$ are frequency-dependent input and output unitary matrices, respectively, * denotes conjugate transpose, and

$$\mathbf{\Lambda}^{(t)}(\omega) = \begin{pmatrix} e^{\frac{1}{2}g_1^{(t)}(\omega)} & & 0 \\ & \ddots & \\ 0 & & e^{\frac{1}{2}g_D^{(t)}(\omega)} \end{pmatrix} \tag{3}$$

is a frequency-dependent real diagonal matrix quantifying MDL, which becomes an identity matrix when MDL is absent.

The input unitary matrix $\mathbf{U}^{(t)}(\omega)$ comprises D column vectors $\mathbf{u}_1(\omega), \mathbf{u}_2(\omega), \ldots, \mathbf{u}_D(\omega)$, each $D \times 1$, which represent mutually orthogonal input eigenmodes of the system, and are eigenvectors of a *phase-conjugate round-trip propagation operator* $\mathbf{M}^{(t)^*}(\omega)\mathbf{M}^{(t)}(\omega)$. Similarly, the output unitary matrix $\mathbf{V}^{(t)}(\omega)$ comprises D column vectors $\mathbf{v}_1(\omega), \mathbf{v}_2(\omega), \ldots, \mathbf{v}_D(\omega)$, each $D \times 1$, which represent mutually orthogonal output eigenmodes of the system, and are eigenvectors of another phase-conjugate round-trip propagation operator $\mathbf{M}^{(t)}(\omega)\mathbf{M}^{(t)^*}(\omega)$. In various fields of physics, the input and output eigenmodes may be referred to as *Schmidt modes* [31], [32]. In the wireless literature, they may be referred to as *spatial channels*, and $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)}(\omega)$ may be called *transmit* and *receive beamforming matrices*, respectively. The input and output eigenmodes represent an optimal basis for MIMO transmission without crosstalk at frequency ω.

The diagonal matrix $\mathbf{\Lambda}^{(t)}(\omega)$ can be described equivalently by a $1 \times D$ frequency-dependent real vector $\mathbf{g}^{(t)}(\omega)$, whose components $g_1^{(t)}(\omega), g_2^{(t)}(\omega), ..., g_D^{(t)}(\omega)$ are the logarithms of the eigenvalues of $\mathbf{M}^{(t)^*}(\omega)\mathbf{M}^{(t)}(\omega)$ or, equivalently, of $\mathbf{M}^{(t)}(\omega)\mathbf{M}^{(t)^*}(\omega)$. These eigenvalues are the squares of the singular values of $\mathbf{M}^{(t)}(\omega)$. Without loss of generality, we assume an ordering $g_1^{(t)}(\omega) \geq g_2^{(t)}(\omega) \geq \cdots \geq g_D^{(t)}(\omega)$. We may refer to $\mathbf{g}^{(t)}(\omega)$ as the *MDL vector* or as the *spatial channel gain vector*, following the wireless literature. The components of $\mathbf{g}^{(t)}(\omega)$ describe the gains between the input and output

eigenmodes, and determine the capacity or performance of an MDM system.

The SVD (2) can be rewritten as

$$\mathbf{M}^{(t)}(\omega) = \sum_{m=1}^{D} e^{\frac{1}{2} g_k^{(t)}(\omega)} \mathbf{v}_m(\omega)\, \mathbf{u}_m^*(\omega). \tag{4}$$

In physical terms, the expansion (4) states that an MDM system is equivalent to a mode converter that maps the mth input eigenmode $\mathbf{u}_m(\omega)$ to the mth output eigenmode $\mathbf{v}_m(\omega)$ with a modal power gain $e^{g_m^{(t)}(\omega)}$ [33]. The expansion (4) simply adds a frequency dependence to the expansion given in [33]. The expansion (4) is helpful in designing an optimal MDM system. In such a system, input data streams are transmitted on input eigenmodes $\mathbf{u}_1(\omega), \mathbf{u}_2(\omega), \ldots, \mathbf{u}_D(\omega)$ and the output data streams are received from output eigenmodes $\mathbf{v}_1(\omega), \mathbf{v}_2(\omega), \ldots, \mathbf{v}_D(\omega)$. Equivalently, in the terminology of wireless systems, input and output data streams are mapped to and from the channel using transmit and receive beamforming matrices $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)^*}(\omega)$, respectively. The data streams at each frequency ω are free of crosstalk, since $\mathbf{V}^{(t)^*}(\omega)\mathbf{M}^{(t)}(\omega)\mathbf{U}^{(t)}(\omega) = \mathbf{\Lambda}^{(t)}(\omega)$ is diagonal [13], [17], [34].

In order to implement this optimal transmission scheme, the receiver must estimate the propagation matrix $\mathbf{M}^{(t)}(\omega)$ at each frequency, compute the SVD (2), and send channel state information (CSI) to the transmitter. This CSI includes the transmit beamforming matrix $\mathbf{U}^{(t)}(\omega)$ and the MDL vector $\mathbf{g}^{(t)}(\omega)$. The transmitter precodes the data using $\mathbf{U}^{(t)}(\omega)$, allocating transmit power and information bits to the spatial channels based on the MDL vector $\mathbf{g}^{(t)}(\omega)$ [17], [21]. In practice, the transmit beamforming matrix $\mathbf{U}^{(t)}(\omega)$ may be estimated imperfectly, and feedback of CSI may be subject to error.

Assuming the components of $\mathbf{g}^{(t)}(\omega)$ are distinct, the SVD (2) yields unique input and output eigenmodes up to multiplicative phase factors of the form $\exp(j\varphi(\omega))$. These phases on the transmit eigenmodes can be adjusted to simplify the description of $\mathbf{U}^{(t)}(\omega)$ for feedback of CSI. While unlikely, in the case that some components of the MDL vector $\mathbf{g}^{(t)}(\omega)$ are degenerated and correspond to more than one linearly independent eigenmode, the description of $\mathbf{U}^{(t)}(\omega)$ can be simplified further.

Feedback of CSI may become impractical if the MMF changes on a time scale shorter than or comparable to the round-trip propagation delay, which can be tens of milliseconds in a long-haul system. When feedback becomes impossible, space-time codes or error-correction codes across the spatial channels can provide diversity [35]–[37]. In such cases, all spatial channels are allocated equal power. In MDM systems, the frequency dependence of $\mathbf{g}^{(t)}(\omega)$ may be exploited to provide frequency diversity [24] by using wideband single-carrier modulation or multicarrier modulation with error-correction coding across frequency.

The model given here, although described in the terminology of MDM systems, is valid for any linear MIMO system. While we have considered a $D \times D$ propagation matrix $\mathbf{M}^{(t)}(\omega)$ in (1), a square matrix is not required, and wireless MIMO systems often use unequal number of transmit and receive antennas. As the major goal of MDM systems is to increase system throughput, a $D \times D$ propagation matrix with D parallel data streams represents the best utilization of available resources. Likewise, an MDM system should be designed so that $\mathbf{M}^{(t)}(\omega)$ is full rank with D nonzero singular values in the SVD (4). Any very small or zero singular values represent a loss of available modes and thus a loss of system throughput.

Aside from MIMO systems, SVD is used in principal component analysis [38], discrete Karhunen–Loève transforms [39], genome data processing [40], channel estimation [41], and many other applications. Highly efficient numerical SVD algorithms are available [30, Sec. 2.5], [42].

B. Schmidt Modes

The input and output unitary matrices $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)}(\omega)$ appearing in (2) and their columns, which represent input and output eigenmodes of a linear MIMO system, may be worthy of more detailed discussion.

In SMF with two polarization modes, where $D = 2$, MDL is called *polarization-dependent loss* (PDL) [43]–[45]. For $D = 2$, the 2×2 unitary matrices $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)}(\omega)$ represent the two orthogonal eigenmodes [46] with maximum and minimum gains, and may be represented as Stokes vectors on the Poincaré sphere, unit-norm Jones vectors, or in other equivalent forms.

In MMF with MDL, where $D > 2$, the eigenmodes represent directions along which gain is an extremum, which may be either a maximum or minimum when approached from different orientations. For $D = 3$, MDL may be visualized in a three-dimensional space using a triaxial ellipsoid with three unequal semi-axes, where the distance from the origin to a point on the surface represents modal gain. The directions of maximal and minimal gain are obvious. The triaxial ellipsoid has a third extremum, which may be either a maximum or minimum, depending on the direction from which it is approached.

In certain fields of physics, including optics and quantum information, the SVD (2) is sometimes called a *Schmidt decomposition* [31], [32], [47], and the eigenmodes described by the columns of $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)}(\omega)$ may be called *Schmidt modes*. Schmidt modes are a useful analytical tool, but their physical interpretation requires some care.

The modes of a laser resonator [48] are eigenmodes of a symmetric, non-Hermitian operator describing forward and backward propagation through the resonator, given in the notation used here by $\mathbf{M}^{(t)T}(\omega)\mathbf{M}^{(t)}(\omega)$, where ^{T} denotes transpose. By contrast, the input Schmidt modes are eigenmodes of the Hermitian operator $\mathbf{M}^{(t)^*}(\omega)\mathbf{M}^{(t)}(\omega)$, which represents forward propagation, phase conjugation, backward propagation, and phase conjugation, and may be called a *phase-conjugate round-trip propagation operator*. The operator $\mathbf{M}^{(t)}(\omega)\mathbf{M}^{(t)^*}(\omega)$, whose eigenmodes are the output Schmidt modes, has an analogous physical interpretation. Either set of Schmidt modes reproduces itself after phase-conjugate round-trip propagation with a scaling given, in logarithmic units, by the components of the MDL vector $g_1^{(t)}(\omega), g_2^{(t)}(\omega), \ldots, g_D^{(t)}(\omega)$.

In a laser resonator, the lasing mode is equivalent to the eigenmode of a round-trip propagation operator $\mathbf{M}^{(t)T}(\omega)\mathbf{M}^{(t)}(\omega)$ that has the largest eigenvalue. The numerical method for

computing resonator modes in [48] is similar to the power method used to find the eigenmode having the largest singular value or eigenvalue [30, Sec. 7.3], which finds application in the Google Pagerank algorithm [49] and beamforming [50]. In an MDM system, the dominant Schmidt mode may be found using a similar power method, and would provide the best performance for transmission of a single data stream.

The condition number of the propagation matrix $\mathbf{M}^{(t)}(\omega)$ is defined as the ratio between the maximum and minimum singular values. In logarithmic units, the condition number equals the maximum MDL difference $g_1^{(t)}(\omega) - g_D^{(t)}(\omega)$. When the condition number is high, MIMO system capacity tends to be reduced, and the MIMO channel is difficult to numerically invert (equalize by zero-forcing).

In wireless MIMO systems, it is difficult or impossible to control the channel properties and thus control the condition number. Wireless MIMO systems often use the condition number as a metric to determine operating mode (choice of modulation and code rate) [51], [52]. By contrast, in MDM systems, MDL can be controlled by careful design of link components, up to physical and economic limits, of course. The design goal for MDM systems is to reduce MDL, such that the propagation matrix $\mathbf{M}^{(t)}(\omega)$ is close to unitary.

In MDM systems, transmission fibers typically have small MDL, and the dominant source of MDL is likely to be from optical amplifiers that compensate for fiber loss [53], [54]. Wavelength-selective switches may also induce MDL, especially at frequencies near the passband edge. In any case, the design goal is to reduce the MDL of each component and thus minimize the MDL of the overall system.

C. Principal Modes

The SVD (2) has a clear physical interpretation and clear implications for MDM system performance. As stated above, a system design goal is to make the propagation operator $\mathbf{M}^{(t)}(\omega)$, given by (1), close to unitary (apart from a multiplicative factor). If $\mathbf{M}^{(t)}(\omega)$ is indeed unitary, the SVD (2) is not uniquely defined, and there are an infinite number of choices for the input and output unitary matrices $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)}(\omega)$. Unitary matrices are generalized rotation matrices, and any arbitrary "rotation" incorporated in $\mathbf{U}^{(t)}(\omega)$ can be compensated by incorporating the inverse "rotation" in $\mathbf{V}^{(t)}(\omega)$.

In the special case of a unitary propagation operator $\mathbf{M}^{(t)}(\omega)$, the input unitary matrix $\mathbf{U}^{(t)}(\omega)$ can be defined uniquely by a further constraint that it be independent of frequency to first order, i.e., $\mathrm{d}\mathbf{U}^{(t)}(\omega)/\mathrm{d}\omega = 0$. With this constraint, the columns of $\mathbf{U}^{(t)}(\omega)$ are *input principal modes*, which are eigenvectors of a *group delay operator* [13], [20], [55], [56]

$$\mathbf{G}(\omega) = j\frac{\mathrm{d}\mathbf{M}^{(t)^*}(\omega)}{\mathrm{d}\omega}\mathbf{M}^{(t)}(\omega). \tag{5}$$

If $\mathbf{M}^{(t)}(\omega)$ is unitary, $\mathbf{G}(\omega)$ is Hermitian with real eigenvalues $\boldsymbol{\tau}^{(t)} = (\tau_1^{(t)}, \tau_2^{(t)}, ..., \tau_D^{(t)})$, representing the group delays of the different principal modes. We assume that the principal mode group delays are ordered as $\tau_1^{(t)} \geq \tau_2^{(t)} \geq \cdots \geq \tau_D^{(t)}$. The *output principal modes* are given by the columns of $\mathbf{V}^{(t)}(\omega)$, which are eigenvectors of a group delay operator $j\mathrm{d}\mathbf{M}^{(t)}(\omega)/\mathrm{d}\omega\mathbf{M}^{(t)^*}(\omega)$ with the same group delays.

We restrict attention to optical signals occupying a narrow bandwidth near frequency ω. If an input principal mode $\mathbf{u}_m$ is launched, the field at the fiber output is described by the corresponding output principal mode $\mathbf{v}_m = \mathbf{M}^{(t)}(\omega)\mathbf{u}_m$. Each of the input principal modes is defined such that if we fix the input field pattern to be $\mathbf{u}_m$ and vary ω slightly, the output field pattern $\mathbf{v}_m$ remains unchanged to first order in ω. The input and output principal modes form the columns of the unitary matrix of $\mathbf{U}^{(t)}$ and $\mathbf{V}^{(t)}$ that are independent of frequency (to first order, and dropping the frequency dependence here). For signals occupying a sufficiently narrow bandwidth near frequency ω, the overall input–output relationship of the MMF can be expressed as:

$$\mathbf{M}^{(t)}(\omega) = \mathbf{V}^{(t)}\boldsymbol{\Lambda}^{(t)}(\omega)\,\mathbf{U}^{(t)^*} \tag{6}$$

where the matrix

$$\boldsymbol{\Lambda}^{(t)}(\omega) = \begin{pmatrix} e^{-j\omega\tau_1^{(t)}} & & 0 \\ & \ddots & \\ 0 & & e^{-j\omega\tau_D^{(t)}} \end{pmatrix} \tag{7}$$

describes propagation of the principal modes. The diagonal matrix (7) represents signal propagation with no crosstalk, and describes differential delay but no distortion, since its phase depends linearly on ω. Many factorizations of the form (6) like (2) are possible mathematically, but this choice is unique in yielding frequency-independent $\mathbf{U}^{(t)}$ and $\mathbf{V}^{(t)}$ (to first order only), and a diagonal $\boldsymbol{\Lambda}^{(t)}(\omega)$ with well-defined group delays. The principal modes in (6) are a special case of the SVD (2).

This description of MD in MMF is consistent with *polarization-mode dispersion* (PMD) in SMF [46], [57], [58], which is a special case with $D = 2$. The PMD operator is a special case of (5), and the two principal states of polarization (PSP) are given by a decomposition like (6). The PSPs have group delays $\tau_1^{(t)} = -\tau_2^{(t)} = \Delta\tau/2$, where $\Delta\tau$ is the differential group delay.

In direct-detection systems in SMF, the PSPs form the basis for optical methods to avoid or compensate PMD [58]. The performance of these methods is limited by higher order PMD [58] and PDL [43]–[45]. Likewise, in direct-detection systems in MMF, the decomposition (6) implies that principal modes may form the basis for optical signal processing methods to avoid or compensate MD and modal crosstalk, which might enable MDM without coherent detection and MIMO digital signal processing. The performance of such methods is expected to be limited by higher-order MD and by MDL [13].

In MDM systems using coherent detection with MIMO signal processing, MD is not fundamentally a performance-limiting factor. The MIMO equalizer must have sufficient complexity to equalize the channel [13], [27], which depends on the maximum group delay spread $\tau_1^{(t)} - \tau_D^{(t)}$ (see Section IV below). It is also imperative that the group delay spread be far smaller than the time constants for temporal variation of the transmission fiber and of laser phase noise.

The group delay operator $\mathbf{G}(\omega)$ (5) can be defined generally for any propagation operator $\mathbf{M}^{(t)}(\omega)$ (1), even one that is not unitary because of MDL [55]. If $\mathbf{M}^{(t)}(\omega)$ is not unitary, the eigenvalues of $\mathbf{G}(\omega)$ (5) are typically complex. While the input principal modes are mutually orthogonal, propagation described by the operator $\mathbf{M}^{(t)}(\omega)$ does not typically map them to orthogonal output principal modes. In general, using principal modes, it is impossible to simultaneously avoid MD and modal crosstalk, potentially impairing MDM systems using direct detection, as noted previously [13], [59].

If the propagation operator $\mathbf{M}^{(t)}(\omega)$ (1) is close to unitary, the group delay operator $\mathbf{G}(\omega)$ is close to Hermitian and has eigenvalues with small imaginary parts. The principal modes become a good approximation to the Schmidt modes. In an MDM system using coherent detection, if the input principal modes are fed back to the transmitter as CSI and used as a transmit basis, then the receiver's MIMO signal processing could be simplified. As noted in Section II-A, however, CSI feedback is not feasible in long-haul systems, and such a feedback is not absolutely reliable.

In a system with MDL, it is mathematically possible for the propagation operator $\mathbf{M}^{(t)}(\omega)$ to have degenerate singular values in (3). Schmidt modes with degenerate singular values form a subspace, and a set of principal modes may be found within that subspace. Below, after discussing mode coupling effects, we explain why the propagation operator $\mathbf{M}^{(t)}(\omega)$ with MDL is unlikely to have degenerate singular values.

In wireless MIMO systems, a group delay operator like (5) can be defined. In a typical wireless system, the channel impulse response resembles a train of impulses, each corresponding to a physical scatterer or reflector [22], [23]. The delay spread is equivalent to the maximum path difference between scatterers. To some extent, an MDM system in MMF is analogous to a wireless MIMO link having one scatterer corresponding to each nondegenerate principal mode. However, the principal modes are mutually orthogonal, and are without obvious analogues in wireless MIMO systems.

III. Mode Coupling

In MMFs, signals propagating in different modes are coupled by random or intentional perturbations. In this section, we discuss mode coupling, including its physical origins, its impact on systems, and power- and field-coupling models used to describe it. We review a multisection field coupling model used to describe propagation in long MMFs. We compare the regimes of weak and strong mode coupling, and describe the scaling of group delay spread and the loss or gain spread in the strong-coupling regime of interest for long-haul MDM systems.

A. Origins and Models

In an idealized fiber, the waveguide modes propagate without cross-coupling such that the propagation operator $\mathbf{M}^{(t)}(\omega)$ is diagonal, assuming those waveguide modes are chosen as bases at both the input and output. In real fibers, perturbations, whether intended or not, can induce coupling between spatial and/or polarization modes [14], [15]. Only coupling between forward-propagating modes is considered here, since it has a dominant effect on the system properties of interest, including MD and MDL.

In MMFs, mode coupling can arise from several unintended sources. These include manufacturing variations causing noncircularity of the core, roughness at the core-cladding boundary, variations in the core radius, or index-profile variations in graded-index fibers. They also include stresses induced by the jacket, or by thermal mismatches between glasses of different compositions. Finally, mode coupling can arise from microbending, macro-bending, or twists.

Mode coupling can have various impacts on transmission systems in MMF [13]. In short-range systems using direct detection, mode coupling is mainly deleterious. In a conventional single-input single-output link, even if a signal is launched into one mode, it becomes coupled into other modes, making MD unavoidable and limiting achievable bit rates in direct-detection links. On the other hand, in plastic MMFs, mode coupling reduces the group delay spread [60], [61] and enables higher bit rates, although it greatly increases loss. If MDM is attempted using direct detection [62]–[64], mode coupling can cause crosstalk between data streams. Such mode coupling must be prevented or compensated by adaptive optical signal processing, since it cannot be fully undone by electrical signal processing after direct detection. By contrast, in long-haul MDM systems using coherent detection, mode coupling is mainly beneficial. It seems virtually impossible to avoid mode coupling in a long-haul system, so full-rank MIMO signal processing becomes necessary [65]. But strong mode coupling minimizes the group delay spread from MD, minimizing the complexity of MIMO signal processing [13], [27]. Similarly, strong mode coupling minimizes the variation of gain or loss from MDL, minimizing any loss of capacity and minimizing the potential for outage [24].

Mode coupling can be described by *field coupling models* or *power coupling models* [13], [66]. Field coupling models describe phase-dependent coupling between complex-valued electric field amplitudes, and are necessary to describe how mode coupling affects the eigenmodes and eigenvalues of certain operators of interest. For example, field coupling models are required in studying the phase-conjugate round-trip operator $\mathbf{M}^{(t)^*}(\omega)\mathbf{M}^{(t)}(\omega)$ or the group delay operator $\mathbf{G}(\omega)$ and their corresponding eigenmodes, the Schmidt modes or principal modes. Thus, field coupling models are essential for a full description of MDM systems. Power coupling models describe coupling between real-valued modal powers by nonnegative real coupling coefficients, a type of diffusion process, and cannot describe changes in eigenmodes and their associated eigenvalues caused by mode coupling. Nevertheless, power coupling models are still useful in giving some qualitative insights into certain aspects of MDM systems.

In field coupling models, a refractive index perturbation inducing mode coupling is often factored to separate its dependence on transverse and longitudinal coordinates. Given a longitudinal dependence $f(z)$, the complex-valued pairwise coupling coefficient between two modal fields is proportional

to [13], [15], [66, eq. 3.4–10]

$$F(\Delta\beta) = \int f(z) e^{-j\Delta\beta z} \mathrm{d}z \tag{8}$$

where the integration is performed over the propagation interval of interest, and $\Delta\beta$ is the difference between the propagation constants of the two modes. The expression (8) can be interpreted as the Fourier transform of $f(z)$ evaluated at spatial frequency $\Delta\beta$. In power coupling models, the longitudinal dependence $f(z)$ is interpreted as a stationary random process, rather than a fully specified function, and the real-valued pairwise coupling coefficient between two modal powers is proportional to $\langle |F(\Delta\beta)|^2 \rangle$, where the brackets $\langle\rangle$ denote an ensemble average [66, eq. 5.2–20].

In typical glass MMFs, most random perturbations have longitudinal power spectra that are low pass. Depending on the autocorrelation function assumed for the perturbations, one finds that $\langle |F(\Delta\beta)|^2 \rangle \propto \Delta\beta^{-4}$ to $\Delta\beta^{-8}$ [15] or $\langle |F(\Delta\beta)|^2 \rangle \propto \Delta\beta^{-2}$ [67] (a constant deterministic coupling coefficient also yields $\langle |F(\Delta\beta)|^2 \rangle \propto \Delta\beta^{-2}$ if not modeled using local normal modes [13, sec. 11.2.3.1], [66, sec 6.3]). As a consequence, typical random perturbations strongly couple modes that are nearly degenerate (small $\Delta\beta$), but only weakly couple modes that are strongly nondegenerate modes (large $\Delta\beta$). Similarly, in typical SMFs, the two nearly degenerate polarization modes ($\Delta\beta \approx 0$) are strongly coupled by random birefringence, which has a low pass power spectrum.

These considerations can be used to explain the evolution of coupling between a set of MDM signals as they propagate along a MMF. This discussion assumes the fiber has small MDL, as is typical of glass fibers, and assumes a basis of ideal waveguide modes.

If the MMF is very short (less than few meters), the propagation matrix $\mathbf{M}^{(t)}(\omega)$ should be close to an identity matrix $\mathbf{I}$, apart from a possible rotation between the input and output coordinate system.

After 5 to 15 meters, the two polarization modes of each spatial mode strongly couple with each other and the MMF enters the polarization coupling state [68], [69]. The propagation matrix $\mathbf{M}^{(t)}(\omega)$ becomes block diagonal, with a sequence of $D/2$ submatrices along the diagonal. Each of these 2×2 unitary submatrices is equivalent to a PMD transfer matrix [46].

After a distance typically less than 100 to 300 meters, all spatial modes within a spatial mode group, which have nearly equal propagation constants, become fully coupled, and the MMF is in the mode-group coupling state [63], [70]. For example, in a graded-index MMF, the gth group has $2g$ modes. The fiber has total of $g(g+1)$ modes up to the gth mode group, including two polarization modes per spatial mode. Each mode group may be considered together as a tubular mode based on the power flow model [14], [62], [63].

In the mode-group coupling state, the propagation matrix $\mathbf{M}^{(t)}(\omega)$ is block diagonal, with submatrices along the diagonal of size $2 \times 2, 4 \times 4, 6 \times 6, \ldots$, each of which is a random unitary matrix. Mode-group division multiplexing assumes the propagation matrix has this block-diagonal structure, with crosstalk within each group but not between groups [62], [63]. Either angle [71] or radially offset [62] launching can launch signals preferentially to specific mode groups. Coupling between groups remains small for distances up to 300 m [72].

After propagation of several km to hundreds of km, if propagation constant differences between mode groups are sufficiently small, all modes fully couple with each other [65], [73], [74]. For typical MMFs with small MDL, the propagation matrix $\mathbf{M}^{(t)}(\omega)$ approaches a $D \times D$ unitary matrix.

B. Multisection Model

The *multisection model* is useful in studying the effect of random inhomogeneity and mode coupling on properties of interest, including modal gains and losses or group delays. Multisection models have been widely used in studying PMD and PDL in SMF [43]–[46]. A MMF link is subdivided into K shorter sections, represented by propagation matrices $\mathbf{M}^{(k)}(\omega), k = 1, \ldots, K$. Propagation through the cascade of K fiber sections is represented by a total propagation matrix $\mathbf{M}^{(t)}(\omega)$, which is the product of the K matrices [13], [20], [21], [27]–[29]

$$\mathbf{M}^{(t)}(\omega) = \mathbf{M}^{(K)}(\omega)\mathbf{M}^{(K-1)}(\omega)\cdots\mathbf{M}^{(2)}(\omega)\mathbf{M}^{(1)}(\omega). \tag{9}$$

Propagation and mode coupling in the kth section is modeled by a product of three $D \times D$ matrices:

$$\mathbf{M}^{(k)}(\omega) = \mathbf{V}^{(k)}\mathbf{\Lambda}^{(k)}(\omega)\mathbf{U}^{(k)^*},\ k = 1, \ldots, K \tag{10}$$

where $\mathbf{\Lambda}^{(k)}(\omega)$ is a frequency-dependent diagonal matrix representing uncoupled propagation in the kth section. The matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ are frequency-independent unitary matrices representing mode coupling at the input and output of the kth section, respectively.

The matrix model (9) and (10) can describe signal propagation through any cascade of linear elements, including fibers, optical amplifiers, optical filters, and modal (de)multiplexers. Difference between various components is mainly reflected in the diagonal matrix $\mathbf{\Lambda}^{(k)}(\omega)$. Including MDL, MD, and mode-dependent chromatic dispersion (MDCD), uncoupled propagation in the kth section is described by a diagonal matrix (11) as shown at the bottom of the page, where $\mathbf{g}^{(k)} = (g_1^{(k)}, g_2^{(k)}, \ldots, g_D^{(k)})$ specifies the uncoupled modal gains, $\boldsymbol{\tau}^{(k)} = (\tau_1^{(k)}, \tau_2^{(k)}, \ldots, \tau_D^{(k)})$ specifies the uncoupled modal group delays, and $\Delta\boldsymbol{\beta}_2^{(k)} = (\Delta\beta_{2,1}^{(k)}, \Delta\beta_{2,2}^{(k)}, \ldots, \Delta\beta_{2,D}^{(k)})$

$$\Lambda^{(k)}(\omega) = \begin{pmatrix} e^{\frac{1}{2}g_1^{(k)} - j\omega\tau_1^{(k)} - \frac{j}{2}\omega^2\Delta\beta_{2,1}L^{(k)}} & & 0 \\ & \ddots & \\ 0 & & e^{\frac{1}{2}g_D^{(k)} - j\omega\tau_D^{(k)} - \frac{j}{2}\omega^2\Delta\beta_{2,D}L^{(k)}} \end{pmatrix} \tag{11}$$

specifies the uncoupled modal group-velocity dispersion variations. The length of the kth section is $L^{(k)}$. In analyzing statistical variations of MDL, MD and MDCD, we set the average values of gain, delay and dispersion to zero: $g_1^{(k)} + g_2^{(k)} + \cdots + g_D^{(k)} = 0, \tau_1^{(k)} + \tau_2^{(k)} + \cdots + \tau_D^{(k)} = 0$ and $\Delta\beta_{2,1}^{(k)} + \Delta\beta_{2,2}^{(k)} + \cdots + \Delta\beta_{2,D}^{(k)} = 0$. If it is desired to include mode-averaged gains, delays and dispersions, we can multiply (11) by a constant factor

$$\exp\left(\frac{1}{2}\bar{g}^{(k)} - j\omega\bar{\tau}^{(k)} - \tfrac{j}{2}\omega^2\bar{\beta}_2^{(k)}L^{(k)}\right) \tag{12}$$

where $\bar{g}^{(k)}, \bar{\tau}^{(k)}$ and $\bar{\beta}_2^{(k)}$ are the mode-averaged gain, group delay and group-velocity dispersion in the kth section.

The number of sections K and the properties of unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}, k = 1, \ldots, K$ are important in determining the regime of mode coupling, and are addressed in this and the following subsection.

For correlation of fields propagating along the link, the unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k-1)}$ defined in (10) are correlated. If two adjacent sections are highly correlated, we have $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)} \approx \mathbf{I}_D$. Combining sections $k-1$ and k is equivalent to multiplying the two diagonal matrices $\mathbf{\Lambda}^{(k)}(\omega)$ and $\mathbf{\Lambda}^{(k-1)}(\omega)$, so both MDL and MD accumulate linearly.

In the polarization-coupled state, the product $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}$ is a block diagonal matrix with 2×2 unitary submatrices along the diagonal, and the product $\mathbf{\Lambda}^{(k)}(\omega)\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}\mathbf{\Lambda}^{(k-1)}(\omega)$ is block diagonal. In the mode group-coupled state, the product $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}$ is block diagonal with submatrices along the diagonal whose sizes correspond to the number of modes within the mode groups.

If the product $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}$ is block diagonal, the product $\mathbf{M}^{(k)}(\omega)\mathbf{M}^{(k-1)}(\omega)$ becomes $\mathbf{V}^{(k)}\mathbf{\Lambda}^{(k,k-1)}(\omega)\mathbf{U}^{(k-1)}$, where $\mathbf{\Lambda}^{(k,k-1)}(\omega)$ is block diagonal. For the case $D = 6$, for example, we may have (13) as shown at the bottom of the page, where $\mathbf{A}_2$ and $\mathbf{A}_4$ are 2×2 and 4×4 matrices, which are approximately unitary, representing the mode groups with 2 and 4 modes, respectively. Because there is a larger group delay difference between mode groups than within each one, the group delays $\tau_i^{(k,k-1)}, i = 1, 2$ are very close to the sums of the corresponding group delays in $\boldsymbol{\tau}^{(k-1)}$ and $\boldsymbol{\tau}^{(k)}$. The modal gains $g_i^{(k,k-1)} i = 1, 2$ are also close to the sums of the corresponding gains in $\mathbf{g}^{(k-1)}$ and $\mathbf{g}^{(k)}$. For all cases in which a MMF is not fully coupled, both MD and MDL increase linearly with the number of sections or the total fiber length.

Physically, two successive sections are independent of each other if there is complete mode coupling between them (for example, if a mode scrambler is inserted in between them). Mathematically, they are independent of each other if $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}$ is statistically the same as a random unitary matrix. With sufficient mode coupling, the product $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}$ will simply be a random unitary matrix.

In Section II-C, we discussed the possibility that the end-to-end propagation matrix (1) or (9) may have MDL but degenerate singular values, such that principal modes exist in the subspaces of Schmidt modes with degenerate singular values. If $\mathbf{\Lambda}^{(k)}(\omega)$ and $\mathbf{\Lambda}^{(k-1)}(\omega)$ have MDL, even with just two degenerate singular values each, the product $\mathbf{\Lambda}^{(k)}(\omega)\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}\mathbf{\Lambda}^{(k-1)}(\omega)$ has nondegenerate singular values if $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}$ is statistically the same as a random unitary matrix. As an example of degenerate singular values, assume that $\mathbf{\Lambda}^{(k)}(\omega) = \mathbf{\Lambda}^{(k-1)}(\omega) = \mathrm{diag}[1, 2, 1, 2]$ and that $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)} = \mathrm{diag}[\mathbf{U}_1, \mathbf{U}_2]$ is a 4×4 block unitary matrix in which $\mathbf{U}_1$ and $\mathbf{U}_2$ are independent 2×2 unitary matrices. The products $\mathrm{diag}[2, 1]\mathbf{U}_1\mathrm{diag}[2, 1]$ and $\mathrm{diag}[2, 1]\mathbf{U}_2\mathrm{diag}[2, 1]$ have independent singular values that depend only on $\mathbf{U}_1$ and $\mathbf{U}_2$. In this example, the product $\mathrm{diag}[2, 1, 2, 1]\mathrm{diag}[\mathbf{U}_1, \mathbf{U}_2]\mathrm{diag}[2, 1, 2, 1]$ is not likely to have degenerate singular values. A general unitary matrix $\mathbf{U}^{(k)}\mathbf{V}^{(k-1)}$ gives more randomness than the more restricted case of $\mathrm{diag}[\mathbf{U}_1, \mathbf{U}_2]$. In practice, in the presence of MDL, with sufficiently random mode coupling between sections, the end-to-end propagation operator is not likely to have degenerate singular values.

C. Weak- and Strong-Coupling Regimes

Modal fields propagating along a MMF are assumed to be strongly correlated over distances less than or equal to a *correlation length*, and weakly correlated over distances far larger than the correlation length. This correlation length is a generalization of the polarization correlation length used in the study of PMD [46], [68], [69]. In the context of the multisection model, the correlation length corresponds to the minimum section length such that successive unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k-1)}$ are mutually uncorrelated.

Different regimes of mode coupling are familiar from the study of randomly coupled birefringence and PMD, and apply as well to mode coupling in MMF [13], [56]. These regimes are especially relevant in studying key effects, such as MD or MDL.

In the *weak-coupling regime*, the correlation length is comparable to, or slightly shorter than, the total system length [13], [57], [75]. In this regime, signal propagation can be modeled using a *small number* of sections K (only 1 or 2), where each section should be slightly longer than the correlation length. In the weak-coupling regime, the spread of eigenvalues describing quantities of interest, such as modal group delays or modal gains, scales *linearly* with the total system length [13], [73], [75], and each coupled eigenmode is a linear combination of a *small number* of uncoupled waveguide modes. If a large number of sections K is used, successive unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k-1)}$ should be highly correlated with each other.

In the *strong-coupling regime*, the correlation length is far shorter than the total system length [13], [57], [58], [75]. In this regime, signal propagation must be modeled using a *large*

$$\mathbf{\Lambda}^{(k,k-1)}(\omega) = \begin{pmatrix} e^{\frac{1}{2}g_1^{(k,k-1)} - j\tau_1^{(k,k-1)}\omega}\mathbf{A}_2(\omega) & 0 \\ 0 & e^{\frac{1}{2}g_2^{(k,k-1)} - j\tau_2^{(k,k-1)}\omega}\mathbf{A}_4(\omega) \end{pmatrix} \tag{13}$$

number of sections K, where each section should be slightly longer than the correlation length. The unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k-1)}$ should be statistically independent. In the strong-coupling regime, the spread of eigenvalues describing quantities of interest, such as modal group delays or modal gains, scales with the *square-root* of the number of sections K or the *square-root* of the total system length and, statistically speaking, each coupled eigenmode is a linear combination of *all* uncoupled modes.

When all K sections are statistically independent of each other, the end-to-end propagation matrix $\mathbf{M}^{(t)}(\omega)$ (1) and (9) is the products of K independent random matrices. In the strong-coupling regime, the number of sections K becomes large, and many properties of the propagation matrix $\mathbf{M}^{(t)}(\omega)$ may be studied based on random matrix theory [76].

Assuming all fiber sections have identical statistical properties, the overall group delay variance is

$$\sigma_{\mathrm{gd}}^2 = K\sigma_\tau^2 \tag{14}$$

where $\sigma_\tau^2 = \langle(\tau_m^{(k)})^2\rangle$ is the group delay variance in each section. Regardless of number of modes, the overall group delay standard deviation (STD) among modes is always given by $\sigma_{\mathrm{gd}} = \sqrt{K}\sigma_\tau$, which increases with the square-root of the number of sections or the square-root of the total system length.

In the strong-coupling regime with many independent sections, the group delay operator (5) becomes a zero-trace (traceless) Gaussian unitary ensemble. The group delay vector $\boldsymbol{\tau}^{(t)} = (\tau_1^{(t)}, \tau_2^{(t)}, ..., \tau_D^{(t)})$ has the same distribution as the eigenvalues of the zero-trace Gaussian unitary ensemble. The statistics of MD may be studied using many results already known for Gaussian unitary ensembles [76]. Section IV summarizes the statistics of strongly coupled MD.

To study MDL in the strong-coupling regime, assuming all K sections have statistically identical MDL, the accumulated MDL is defined as

$$\xi^2 = K\sigma_g^2 \tag{15}$$

where $\sigma_g^2 = \langle(g_m^{(k)})^2\rangle$ is the variance of the uncoupled MDL in each section. This accumulated MDL $\xi = \sqrt{K}\sigma_g$ increases with the square-root of number of sections or the square-root of the total system length.

As explained above, propagation matrices $\mathbf{M}^{(t)}(\omega)$ in well-designed systems should be close to unitary, i.e., they have small MDL. In the small-MDL regime, the distribution of the overall MDL (measured in units of the logarithm of power gain or decibels) is well-approximated by the eigenvalue distribution of a zero-trace Gaussian unitary ensemble. In the large-MDL regime, MDL has the same distribution as the eigenvalues of a Rician random matrix or a nonzero-mean Gaussian unitary ensemble. Furthermore, the STD of the overall MDL σ_{mdl} depends solely on the accumulated MDL ξ via [29]

$$\sigma_{\mathrm{mdl}} = \xi\sqrt{1 + \frac{\xi^2}{12(1-D^{-2})}}. \tag{16}$$

In (16), both σ_{mdl} and ξ are measured in units of the logarithm of power gain. Quantities measured in units of log power gain can be converted to decibels by multiplying by $\gamma = 10/\ln 10 \approx 4.34$, i.e., σ_{mdl} (dB) $= \gamma\sigma_{\mathrm{mdl}}$ (log power gain). Unless noted otherwise, all expressions in this paper assume that both ξ and σ_{mdl} are expressed in log-power-gain units.

The statistics of MDL can be studied using some known results on the products of random matrices [21], [29], [77]. The statistics of strongly coupled MDL are summarized in Section V-A.

Both MD and MDL become frequency-dependent, which can be understood to arise from the different modal group delays appearing in the exponents in (11). In the strong-coupling regime, both the MD and the MDL become correlated with a correlation bandwidth of the order of $1/\sigma_{\mathrm{gd}}$. Frequency-dependent MD can be described as *higher-order MD*, which can has been studied in [24], [78]. Higher-order MD can limit the effectiveness of MD compensation or avoidance using any frequency-independent optical device [13]. Frequency-dependent MDL leads to *frequency diversity* [24], which averages different random realizations of MDL over the bandwidth of a signal, reducing capacity fluctuations and the associated outage probability. Frequency diversity is discussed in Section V-B below.

IV. Modal Dispersion

While MD impairs traditional MMF transmission systems using direct detection, it has no fundamental effect on performance in MDM systems using coherent detection and MIMO digital signal processing. Nevertheless, the group delay spread from MD affects the complexity of MIMO signal processing.

The principal modes associated with MD have a clear physical interpretation. MD can be described easily in the multisection or matrix product model (9), and derivation of its statistics in the strong-coupling regime is straightforward. For these reasons, MD is discussed here, and MDL is deferred to the following section.

A. Concatenation Rule and Group Delay Statistics

For simplicity, we neglect mode-averaged loss and MDL, so all K matrices in the product (9) are unitary, and the overall propagation operator $\mathbf{M}^{(t)}(\omega)$ is unitary. We further assume the strong-coupling regime, so all K matrices in (9) are independent. If the product (9) is substituted into group delay operator $\mathbf{G}(\omega)$ given by (5), $\mathbf{G}(\omega)$ is the summation of K independent Hermitian matrices:

$$\begin{aligned}\mathbf{G} = {} & j\frac{d\mathbf{M}^{(1)^*}}{d\omega}\mathbf{M}^{(1)} + \mathbf{M}^{(1)^*}\left(j\frac{d\mathbf{M}^{(2)^*}}{d\omega}\mathbf{M}^{(2)}\right)\mathbf{M}^{(1)} \\ & + \mathbf{M}^{(1)^*}\mathbf{M}^{(2)^*}\left(j\frac{d\mathbf{M}^{(3)^*}}{d\omega}\mathbf{M}^{(3)}\right)\mathbf{M}^{(2)}\mathbf{M}^{(1)} + \cdots\end{aligned} \tag{17}$$

In (17) and throughout this section, we suppress the frequency dependence to simplify notation.

Before proceeding to derive the group delay statistics, we pause to observe that (17) represents the *concatenation rule*

of MD for MMF. On the right-hand side, the first term is the group delay operator of the first section, the parenthesis in the second term is the group delay operator of the second section, and so on. The total group delay operator is the sum of these individual group delay operators, each transformed by unitary similarity. The second section is transformed by the unitary matrix $\mathbf{M}^{(1)}$, the third section is transformed by the unitary matrix $\mathbf{M}^{(2)}\mathbf{M}^{(1)}$, and so on. This concatenation rule holds true even if the individual sections are not statistically independent or are deterministic.

Returning to the derivation of the statistics of strongly coupled MD, if all K matrices $\mathbf{M}^{(k)}(\omega)$, $k = 1, \ldots, K$, are statistically identical and independent, all matrices in the summation (17) are independent and statistically the same as the first term

$$\mathbf{U}^{(1)}\mathrm{diag}\left[\boldsymbol{\tau}^{(1)}\right]\mathbf{U}^{(1)^*}, \tag{18}$$

where $\mathrm{diag}[\boldsymbol{\tau}^{(1)}]$ is a diagonal matrix with diagonal elements from the vector $\boldsymbol{\tau}^{(1)}$. The matrix $\mathbf{U}^{(1)}\mathrm{diag}[\boldsymbol{\tau}^{(1)}]\mathbf{U}^{(1)^*}$ has zero trace because $\mathrm{diag}[\boldsymbol{\tau}^{(1)}]$ is zero trace. The first term of (17) gives (18). The second term of (17) is $\mathbf{M}^{(1)^*}\mathbf{U}^{(2)}\mathrm{diag}[\boldsymbol{\tau}^{(2)}]\mathbf{U}^{(2)^*}\mathbf{M}^{(1)}$, which is statistically the same as (18) because $\mathrm{diag}[\boldsymbol{\tau}^{(2)}]$ and $\mathbf{M}^{(1)^*}\mathbf{U}^{(2)}$ are statistically the same as $\mathrm{diag}[\boldsymbol{\tau}^{(1)}]$ and $\mathbf{U}^{(1)}$, respectively. Following the same logic, all terms in (17) are statistically the same as (18).

Based on Central Limit Theorem, the summation of many independent identically distributed random variables has a Gaussian distribution. All elements of $\mathbf{G}$ given by (17) are Gaussian distributed. Because all K matrices in the summation (17) have zero trace, group delay operator $\mathbf{G}$ (17) has zero trace. Here, we will evaluate the variance of the elements of $\mathbf{G}$ (17).

As in [79]–[81], averaging over the elements of a random unitary matrix $\mathbf{U}$ gives

$$\begin{aligned}\langle u_{kl}u^*_{mn}u_{ab}u^*_{cd}\rangle &= \frac{\delta_{mk}\delta_{nl}\delta_{ca}\delta_{db}+\delta_{ma}\delta_{nb}\delta_{ck}\delta_{dl}}{D^2-1}\\ &\quad-\frac{\delta_{ma}\delta_{nl}\delta_{ck}\delta_{db}+\delta_{mk}\delta_{nb}\delta_{ca}\delta_{dl}}{D(D^2-1)}\end{aligned} \tag{19}$$

where u_{kl} are the elements of the random unitary matric $\mathbf{U}$, δ_{kl} is the Kronecker delta function, which equals unity if $k = l$ and zero otherwise. To find the variance of the elements $\mathbf{U}^{(1)}\mathrm{diag}[\boldsymbol{\tau}^{(1)}]\mathbf{U}^{(1)^*}$ requires taking into account D^4 different combinations in the form (19). After some algebra, the elements of $\mathbf{U}^{(1)}\mathrm{diag}[\boldsymbol{\tau}^{(1)}]\mathbf{U}^{(1)^*}$ are found to have a variance of $\sigma_\tau^2 D/(D+1)/(D+\delta_{kl}-1)$. The elements of $\mathbf{U}^{(1)}\mathrm{diag}[\boldsymbol{\tau}^{(1)}]\mathbf{U}^{(1)^*}$ are uncorrelated with each other. The elements of $\mathbf{G}$ (17) have variance

$$\frac{KD\sigma_\tau^2}{(D+1)(D+\delta_{kl}-1)}. \tag{20}$$

The group delay variance is the same as (14) and given by

$$\sigma_{\mathrm{gd}}^2 = \tfrac{1}{D}\sum_m\left\langle(\tau_m^{(t)})^2\right\rangle = \tfrac{1}{D}\left\langle\mathrm{tr}\left[\mathbf{G}\mathbf{G}^*\right]\right\rangle = K\sigma_\tau^2 \tag{21}$$

where $\mathrm{tr}[\cdot]$ denote the trace of a matrix.

If the zero-trace Gaussian unitary ensemble $\mathbf{G}$ is studied by a Gaussian unitary ensemble $\mathbf{A}$ via $\mathbf{G} = \mathbf{A} - \mathrm{tr}(\mathbf{A})\mathbf{I}/D$, the Gaussian unitary ensemble $\mathbf{A}$ without trace constraint should have a variance of $DK\sigma_\tau^2/(D^2-1)$ or $D\sigma_{\mathrm{gd}}^2/(D^2-1)$. Comparing Gaussian unitary ensembles with and without zero-trace constraint, for the same overall variance as represented by $\langle\mathrm{tr}[\mathbf{A}\mathbf{A}^*]\rangle$ and $\langle\mathrm{tr}[\mathbf{G}\mathbf{G}^*]\rangle$, the off-diagonal elements of $\mathbf{G}$ are a factor

$$\frac{1}{1-D^{-2}} \tag{22}$$

larger than those of $\mathbf{A}$, while the diagonal elements of $\mathbf{G}$ are a factor

$$\frac{1}{1+D^{-1}}. \tag{23}$$

smaller than those of $\mathbf{A}$.

As derived in [13], [20], in normalized form, the eigenvalue distribution for a zero-trace Gaussian unitary ensemble is

$$\begin{aligned}f_D(x) &= \frac{\exp\left(-\frac{D}{D-1}x^2\right)}{\sqrt{\pi D(D-1)}}\\ &\quad\times\sum_{n=0}^{D-1}\frac{1}{2^n n!}H_n^2\left(\frac{-t}{2\sqrt{D-1}}\right)\Bigg|_{t^k\leftarrow H_k\left(\frac{Dx}{\sqrt{D-1}}\right)}\end{aligned} \tag{24}$$

where the summation gives a $2(D\text{-}1)$-degree polynomial in t. In (24), the power t^k is algebraically substituted by the Hermite polynomial $H_k(Dx/\sqrt{D-1})$. The variance of the normalized probability density (24) is $\frac{1}{2}(D - D^{-1})$. Scaling it to have the same variance as (14) yields the probability density function of the strongly coupled group delay:

$$p_D(x) = \sqrt{\frac{D^2-1}{2\sigma_{\mathrm{gd}}^2 D}}f_D\left(\frac{\sqrt{D^2-1}x}{\sigma_{\mathrm{gd}}\sqrt{2D}}\right). \tag{25}$$

Fig. 1 shows the probability density of group delay, normalized by the STD of group delay, for MMFs with different number of modes. As the number of modes increases, the probability density approaches the well-known Wigner semicircle distribution, denoted by a dashed line in Fig. 1. Only the positive-delay

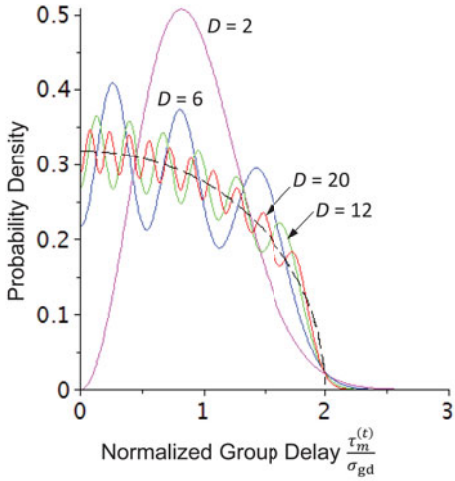

Fig. 1. Probability density of strongly coupled group delay, normalized by the STD of group delay, in MMFs with various numbers of modes. The semicircle distribution is shown as a dashed curve.

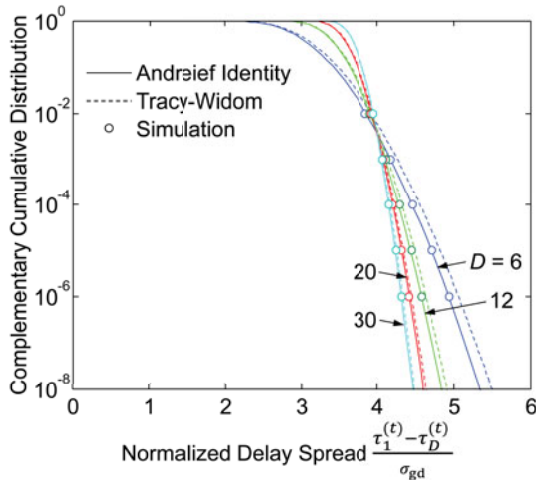

Fig. 2. Complementary cumulative distribution of group delay spread, normalized by the STD of group delay, in MMFs with different numbers of modes.

side is shown in Fig. 1, since all the densities are even about the origin. The number of peaks in each probability density is the same as number of modes D.

B. Delay Spread and Signal Processing Complexity

In an MDM system using digital MIMO equalization, in order to fully compensate MD and modal crosstalk, the equalizer must span a temporal memory at least as long as the system group delay spread $\tau_1^{(t)} - \tau_D^{(t)}$. Here we focus on computing the required equalizer memory length. Detailed studied [13], [27] have addressed the computational complexity and hardware complexity of time-domain or frequency-domain MIMO equalizers, taking account of the memory length and the number of modes D.

From Fig. 1, we expect the group delay spread to approach $4\sigma_{\mathrm{gd}}$ as the number of modes D becomes large. For finite D, the distribution is known for the difference between two group delays $\tau_m^{(t)} - \tau_n^{(t)}$ with random selection of m and n [82]. However, only numerical results are available for the difference between the maximum and minimum group delays $\tau_1^{(t)} - \tau_D^{(t)}$. Fig. 2 shows the complementary cumulative distribution function (CCDF) $\Pr((\tau_1^{(t)} - \tau_D^{(t)})/\sigma_{\mathrm{gd}} > x)$ for MMFs with $D =$ 6, 12, 20, and 30 modes. CCDFs calculated numerically using the Andréief identity [83]–[85] are in excellent agreement with numerical simulations from [27]. For a large number of modes D, the maximum and minimum group delays follow the well-known Tracy–Widom distribution [86], [87] and are independent of each other [88], [89]. Fig. 2 shows CCDFs approximated based on the difference between two independent Tracy-Widom random variables. The Tracy-Widom approximation always overestimates the CCDF, but is sufficiently accurate for most engineering purposes for $D \geq 12$.

Following [27], given a value of D and a probability p, we define $u_D(p)$ to be the value of x such that $\Pr((\tau_1^{(t)} - \tau_D^{(t)})/\sigma_{\mathrm{gd}} > x) = p$. An equalizer memory length of $u_D(p)\sigma_{\mathrm{gd}}$ is sufficient to span the channel memory with probability $1 - p$. For the values of D shown in Fig. 2 and for p of order 10^{-4} to 10^{-6}, $u_D(p)$ is of order 4 to 5, depending on the number of modes D.

The delay spread was studied in [90], which generalized the concept of Stokes space to $D > 2$ and showed that in the strong-coupling regime, the distribution of the delay spread $\tau_1^{(t)} - \tau_D^{(t)}$ is well-approximated by a chi distribution. This is justified heuristically by the observation that the square-root of $\sum_m (\tau_m^{(t)})^2 = \mathrm{tr}[\mathbf{G}\mathbf{G}^*]$ is rigorously chi-square distributed [90].

In practice, system end-to-end group delay spread may be minimized by designing fibers with minimal uncoupled delay spread, and relying on mode coupling to further reduce the delay spread. First proposed in [75], the reduction of delay spread by mode coupling is well-known in the study of PMD in SMF [57], [58] and plastic MMF [60], [61], and has been verified by simulation in [56], [91], [92] for MDM systems. The delay spread of MMF may possibly be reduced by some generalization [93] of spinning, which is used in minimizing PMD in SMF [94], [95].

In the study of MD, the group delay operator is a Gaussian random matrix. However, the overall propagation matrix (1) or (9) is not likely to be Gaussian-distributed, unlike channel models for wireless MIMO systems [17]–[19].

V. Mode-Dependent Gains and Losses

Unlike MD, MDL can fundamentally degrade the performance of an MDM system using coherent detection. Moreover, MDL can make MDM channels ill-conditioned, creating difficulties for digital MIMO equalization.

The Schmidt modes associated with MDL are not amenable to a simple physical interpretation, in contrast to the principal modes for MD. The statistics of MDL are more difficult to study than those of MD. In the strong-coupling regime, the MDL operator is modeled as the product of many random matrices (9), whose statistics are difficult to derive [77], [96]. By contrast, the MD operator can be modeled as the summation of many random matrices (17). Matrix multiplication is noncommutative, whereas matrix summation is commutative.

A. Mode-Dependent Loss/Gain Statistics

We begin by studying MDL at a single frequency, suppressing the frequency dependence in (11). MDL from optical amplifiers typically dominates over that from transmission fibers, so the number of sections K can be taken to equal the number of amplifiers. The propagation operator for each section becomes

$$\mathbf{M}^{(k)} = \mathbf{V}^{(k)} \mathrm{diag}\left[\exp(\tfrac{1}{2}\mathbf{g}^{(k)})\right] \mathbf{U}^{(k)^*}. \tag{26}$$

In each section, the effect of MMF spans before or after the optical amplifiers can be included in either $\mathbf{U}^{(k)}$ or $\mathbf{V}^{(k)}$. The end-to-end MDL $\mathbf{g}^{(t)}$ is the same as the $\mathbf{g}^{(t)}(\omega)$ introduced in Section II-A, but with the frequency dependence suppressed.

In the strong-coupling regime, the MDL $\mathbf{g}^{(t)}$ (expressed in logarithmic or decibel units) has the same statistical properties

as the eigenvalues of the summation of two matrices [29]

$$\xi\mathbf{G} + \frac{D}{2(D+1)}\xi^2\mathbf{F} \tag{27}$$

where $\xi = \sqrt{K}\sigma_g$ is the accumulated MDL defined in (15), $\mathbf{G}$ is a zero-trace Gaussian unitary ensemble similar to that used to describe MD in Section IV, except it has unit eigenvalue variance, and $\mathbf{F}$ is a deterministic uniform matrix. The uniform matrix $\mathbf{F}$ has its eigenvalues deterministically and uniformly distributed between -1 and $+1$. As an example for $D = 2$, we can choose $\mathbf{F} = \text{diag}[1, -1]$ or another unitary similar matrix having the same eigenvalues. The factor $D/2/(D + 1)$ in the second term of (27) is between 1/3 for $D = 2$ and 1/2 for $D \to \infty$.

Like the group delay matrix (17), the matrix (27) is an Hermitian Gaussian random matrix with nonzero-mean diagonal elements. In wireless communications, a random channel in the form (27) is called a Rician MIMO channel [51], [97]. Because $\mathbf{g}^{(t)}$ is measured in logarithmic or decibel units, the MDL channel is a log-Rician MIMO channel.

From the theory of MD given by (17), the summation (27) represents the concatenation of two MMFs: the first with strong mode coupling represented by $\xi\mathbf{G}$ and the other with deterministic and uniform MD represented by the scaling of the matrix $\mathbf{F} = \text{diag}[1, 1 - 2/(D - 1), \ldots, -1]$.

Because matrix multiplication is not commutative, i.e., $\mathbf{AB}$ is not generally equal to $\mathbf{BA}$, even for square matrices, the logarithm of the product of two matrices, $\log\mathbf{AB}$, is not equal to the sum of $\log\mathbf{A}$ and $\text{Log}\mathbf{B}$. The product of positive-definite random matrices (those with positive eigenvalues) does not generally have its central limit as the exponent of a Gaussian unitary ensemble.

For any matrix $\mathbf{X}$ and a very small number δ, we have $\log(\mathbf{I} + \delta\mathbf{X}) \approx \delta\mathbf{X}$, where $\mathbf{I} + \delta\mathbf{X}$ is intended to describe a matrix $\mathbf{M}^{(k)}$ when the gain vector $\mathbf{g}^{(k)}$ has small norm. If both matrices $\mathbf{A}$ and $\mathbf{B}$ are positive-definite and both $\log\mathbf{A}$ and $\log\mathbf{B}$ are small, $\log\mathbf{AB} \approx \log\mathbf{A} + \log\mathbf{B}$. As an approximation, the product of positive-definite random matrices with small logarithm has a central limit as the exponential of a Gaussian ensemble. When applied to the overall product matrix $\mathbf{M}^{(t)} = \mathbf{M}^{(K)} \cdots \mathbf{M}^{(2)}\mathbf{M}^{(1)}$, if all gain vectors $\mathbf{g}^{(k)}$ are small, the matrix $\mathbf{M}^{(t)}$ is the exponential of the Gaussian ensemble, which yields the first term of (27) without the second term. The nonlinearity in (27) relates to the second-order term in the approximation $\log(\mathbf{I} + \delta\mathbf{X}) \approx \delta\mathbf{X} - \frac{1}{2}\delta^2\mathbf{X}^2$. The factor of 1/2 in this approximation is difficult to relate to the factor between 1/2 and 1/3 in (27), however. In any case, the second term of (27) is very small when ξ is far less than unity.

When the propagation operators for the sections $\mathbf{M}^{(k)}$, given by (26), have independent and identically statistics, the second term of (27) is the Lyapunov exponent [98]

$$\lim_{K\to\infty}\frac{1}{K}\log\mathbf{\Lambda}^{(t)} = \frac{D\sigma_g^2}{2(D+1)}\mathbf{F} \tag{28}$$

where $\mathbf{\Lambda}^{(t)}$, appearing in (3) but with frequency dependence suppressed here, is obtained by the SVD (2). If we express the propagation operators as $\mathbf{M}^{(k)} = \exp(\frac{1}{2}\mathbf{C}^{(k)})$ with $\mathbf{C}^{(k)} = \mathbf{V}^{(k)}\text{diag}[\mathbf{g}^{(k)}]\mathbf{U}^{(k)^*}$, the statistics of $\mathbf{C}^{(k)}$ are the same as those of (18) in the study of MD, where the off-diagonal elements have larger variance than the diagonal elements. Using the theory of [98], the Lyapunov exponent (28) can be derived[1].

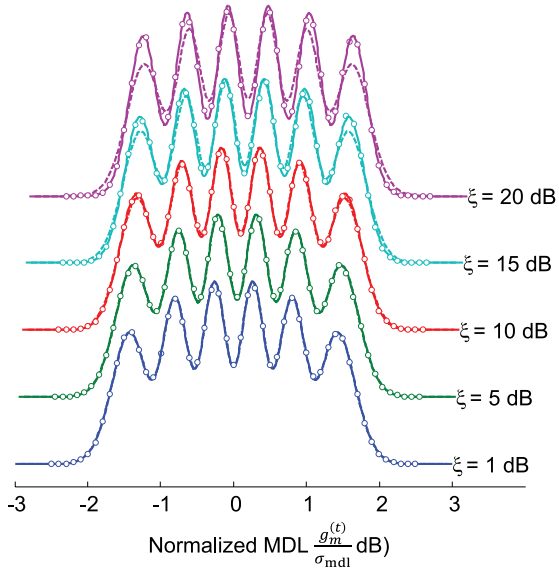

Fig. 3. Probability density of strongly coupled MDL, normalized by the STD of overall MDL, for $D = 6$ modes for various values of the accumulated MDL $\xi = \sqrt{K}\sigma_g$. The exact distribution (27) is shown by solid curves. The approximation based on a Gaussian unitary ensemble is shown by dashed curves. Numerical simulations are shown as markers. The y-axis is in arbitrary linear units, with different curves shifted to improve visibility.

The overall MDL variance σ^2_{mdl}, which is the square of (16), can be derived as the sum of the eigenvalue variances of the first term of (27), which gives ξ^2, and second term of (27), which gives $\xi^4/12/(1 - D^{-2})$. The scale-up factor $1/(1 - D^{-2})$ is the same as (22), evidently due to the zero-trace constraint. The nonlinear term in (16) decreases from 1/9 to 1/12, a factor of 3/4, with an increase from $D = 2$ to $D \to \infty$. In practice, this factor reduces to 1/12 rapidly with an increase in the number of modes.

The second term of (27) can also be approximated as a zero-trace Gaussian unitary ensemble, especially for systems with small MDL of $\xi \leq 10$ dB. Under this assumption, MDL has the same distribution as the zero-trace Gaussian unitary ensemble with STD given by (16). This approximation reduces the analysis of MDL statistics to the analysis of MD statistics given in Section IV. With minor modifications, all results in Section IV and those in [76] are applicable to MDL.

Fig. 3 shows the probability density of MDL normalized by the STD of overall MDL for $D = 6$ modes for various values

[1]Model d or Eq. (15) of [98] shows that the Lyapunov exponent is uniformly distributed with a maximum given by $(D-1)\sigma^2$ where the σ^2 is the variance of the elements of the matrix, using the notation here, $\mathbf{C}_k$. However, if the diagonal and off-diagonal elements have different variances, the step to derive Eq. (15) of shows that σ^2 is the variance of the off-diagonal elements. Equivalently, from Sec. IV, $\sigma^2 = \sigma_g^2 D/(D + 1)/(D - 1)$. The maximum $(D-1)\sigma^2$ becomes $\sigma_g^2 D/(D + 1)$ in (28). The additional factor of $\frac{1}{2}$ comes directly from the factor of $\frac{1}{2}$ in $\mathbf{M}^{(k)} = \exp(\frac{1}{2}\mathbf{C}^{(k)})$.

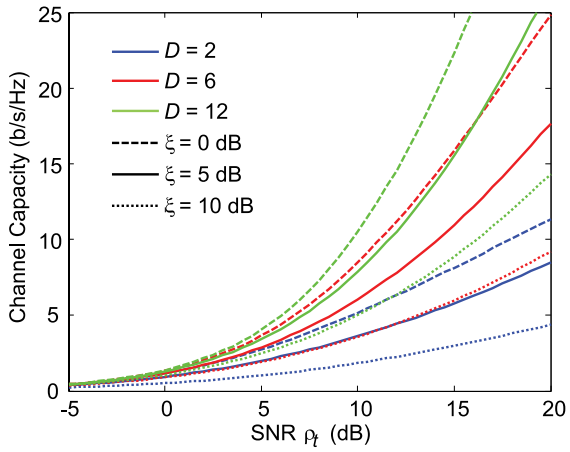

Fig. 4. Average channel capacity of MDM systems without CSI as a function of overall SNR ρ_t. Different numbers of modes D and different values of accumulated MDL ξ are represented by different line colors and styles, respectively.

of the accumulated MDL $\xi = \sqrt{K}\sigma_g$. Fig. 3 compares the exact model (16) as in [29] and the zero-trace Gaussian unitary ensemble approximation of [21] to numerical simulations of a multisection model. The exact model agrees well with simulations, whereas the zero-trace Gaussian unitary ensemble approximation is valid when $\xi \leq 13$ dB [21].

In an MDM system, the condition number of a MIMO channel (expressed in logarithmic units or dB) equals the maximum MDL difference $g_1^{(t)} - g_D^{(t)}$. The Gaussian unitary ensemble approximation can be used to show that the maximum MDL difference $g_1^{(t)} - g_D^{(t)}$ is 4 to 5 times σ_{mdl}, depending on the number of modes D, similar to Fig. 2. To limit the maximum MDL difference to within 30 dB (condition number not exceeding 1000), the accumulated MDL ξ must not exceed 7 dB.

B. System Capacity and Frequency Diversity

MDL fundamentally limits the performance of long-haul MDM systems using coherent detection. MDL and its impact on channel capacity were studied in [13], [21], [24], [28], [92]. The effectiveness of mode coupling in reducing MDL and increasing channel capacity were confirmed in [92], [99]–[101] by simulation.

Due to the difficulty of feeding back reliable CSI to the transmitter, a system must typically allocate equal power to each mode. At a single frequency, the average channel capacity without CSI is [21]

$$C = \left\langle \sum_{m=1}^{D} \log_2 \left[1 + \frac{\chi}{D} \exp\left(g_m^{(t)} \right) \right] \right\rangle \tag{29}$$

where χ is ratio of the total transmitted signal power in all modes to the received noise power in one mode. The signal-to-noise ratio (SNR) is defined as the ratio of total received signal power in all modes to the received noise power in one mode, and is equal to $\rho_t = \chi / \langle \exp g_m^{(t)} \rangle$. Fig. 4 shows the average channel capacity for MDM links with $D = 2$, 6, and 12 modes, with accumulated MDL values $\xi = 0$, 5, and 10 dB. When CSI is not available, the average channel capacity increases monotonically with SNR but decreases with MDL. At low SNRs, the average channel capacity is proportional to the total power and almost independent of number of modes and MDL. At high SNRs, the average channel capacity is reduced by MDL. At $\rho_t = 20$ dB and MDL $\xi = 10$ dB, the average channel capacity is reduced about 40% by MDL, independent of the number of modes.

The instantaneous channel capacity is a random variable that depends on the instantaneous realization of the MDL vector $\mathbf{g}^{(t)}(\omega)$. The system may experience outage if the instantaneous channel capacity drops below a certain value. In general, the MDL vector $\mathbf{g}^{(t)}(\omega)$ is frequency-dependent, so the instantaneous channel capacity is also frequency-dependent. If an MDM system transmits signals occupying a sufficiently wide bandwidth, the MDL vector and the channel capacity will be averaged over frequency, an effect called *frequency diversity*. In a system with sufficient frequency diversity, the instantaneous channel capacity approaches the average channel capacity (29) and the outage probability approaches zero, which are consequences of the law of large numbers.

Numerical results show that the correlation bandwidth of the channel capacity is approximately equal to the reciprocal of the STD of the group delay, $1/\sigma_{\mathrm{gd}}$ [24]. As shown in Section IV-B, the peak-to-peak group delay spread is 4 to 5 times the STD of group delay σ_{gd}. If an MDM system is designed with a group delay spread spanning several hundred symbol intervals [27], the correlation bandwidth of channel capacity is of the order of 1% of the signal bandwidth, providing sufficient frequency diversity to essentially eliminate outage [24], assuming realistically small values of accumulated MDL. Such a delay spread is acceptable from the standpoint of signal processing complexity [13], [27].

In MDM systems with a very small number of amplifiers (e.g., 1 or 2), because of MDL, the noises in different modes can be highly correlated, and may have different variances [28]. In systems using a large number of amplifiers (e.g., in long-haul systems), the noises in different modes become independent and identically distributed [21], i.e., *spatially white*, as a consequence of the law of large numbers. This greatly simplifies system analysis in the linear regime, allowing all noise to be added right before the receiver, and is assumed in (29) and Fig. 4.

VI. Conclusion

MIMO channels in MDM systems exhibit many similarities to wireless MIMO channels. MDL, arising mainly from optical amplifiers, is analogous to spatial channel variations due to multipath fading. By causing fluctuations in gains of different spatial channels, it can reduce average capacity and cause outage in narrowband systems. Likewise, it can increase the MIMO channel condition number, complicating equalization. MD, arising from transmission fibers, is analogous to multipath delay spread. It affects the complexity of MIMO equalization, but has no fundamental impact on performance.

Mode coupling is mainly beneficial in long-haul MDM systems. In the strong-coupling regime, an end-to-end system can be modeled as the concatenation of many uncorrelated sections. Strong mode coupling reduces the variance of modal group delays from MD, minimizing signal processing complexity.

Likewise, it minimizes the variance of modal gains/losses from MDL, improving system performance. In wideband systems with MD, strong coupling leads to frequency diversity, which reduces the probability that the outage channel capacity differs significantly from the average channel capacity.

In the strong-coupling regime, the statistics of MD and MDL are the same as the eigenvalue distribution of zero-trace Hermitian Gaussian random matrices. Modal group delays can be described by zero-mean Gaussian random matrices. Modal gains and losses (in logarithmic units) have the same statistics as the eigenvalue distribution of Gaussian random matrices with nonzero-mean diagonal elements. Alternatively, MDL has the same statistics as the concatenation of two MMF sections: the first is strongly coupled and the second is uniformly distributed. In the low-MDL regime, MDL statistics can be approximated as the eigenvalue distribution of a zero-mean zero-trace Gaussian random matrix.

ACKNOWLEDGMENT

We are grateful to C. J. McKinstrie for calling our attention to applications of Schmidt modes in various fields of physics.

REFERENCES

[1] K. C. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," *Proc. IEE*, vol. 113, no. 7, pp. 1151–1158, 1966.

[2] T. Li, "Advances in optical fiber communications: An historical perspective," *IEEE J. Sel. Areas Commun.*, vol. 1, no. 3, pp. 356–372, Apr. 1983.

[3] A. H. Gnauck, R. W. Tkach, A. R. Chraplyvy, and T. Li, "High-capacity optical transmission systems," *J. Lightw. Technol.*, vol. 29, no. 9, pp. 1032–1045, 2008.

[4] K. Roberts, M. O'Sullivan, K.-T. Wu, H. Sun, A. Awadalla, D. J. Krause, and C. Laperle, "Performance of dual-polarization QPSK for optical transport systems," *J. Lightw. Technol*, vol. 27, no. 16, pp. 3546–3559, 2009.

[5] E. Ip, P. Ji, E. Mateo, Y.-K. Huang, L. Xu, D. Qian, N. Bai, and T. Wang, "100G and beyond transmission technologies for evolving optical networks and relevant physical-layer issues," *Proc. IEEE*, vol. 100, no. 5, pp. 1065–1078, 2012.

[6] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, pp. 1027–1030, Jun. 28, 2001.

[7] 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, 2010.

[8] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," *J. Lightw. Technol.*, vol. 28, no. 6, pp. 939–951, 2010.

[9] A. Al Amin, A. Li, S. Chen, X. Chen, G. Gao, and W. Shieh, "Dual-LP_{11} mode 4×4 MIMO-OFDM transmission over a two-mode fiber," *Opt. Exp.*, vol. 19, no. 17, pp. 16672–16678, 2011.

[10] C. Koebele, M. Salsi, D. Sperti, P. Tran, P. Brindel, H. Margoyan, S. Bigo, A. Boutin, F. Verluise, P. Sillard, M. Bigot-Astruc, L. Provost, F. Cerou, and G. Charlet, "Two mode transmission at 2×100 Gb/s, over 40 km-long prototype few-mode fiber, using LCOS based mode multiplexer and demultiplexer," *Opt. Exp.*, vol. 19, no. 17, pp. 16593–16600, 2011.

[11] R. J. Essiambre, R. Ryf, N. K. Fontaine, and S. Randel, "Space-division multiplexing in multimode and multicore fibers for high-capacity optical communications," *IEEE Photonics J.*, vol. 5, no. 2, p. 071307, 2013.

[12] D. J. Richardson, J. M. Fini, and L E. Nelson, "Space division multiplexing in optical fibres," *Nature Photon.*, vol. 7, pp. 354–362, 2013.

[13] K.-P. Ho and J. M. Kahn, "Mode coupling and its impact on spatially multiplexed systems," in *Optical Fiber Telecommunications VI, B*, I. P. Kaminow, T. Li, and A. E. Willner, Eds. Oxford, U.K.: Academic, 2013, ch.11, pp. 491–568.

[14] D. Gloge, "Optical power flow in multimode fiber," *Bell Syst. Tech. J.*, vol. 51, no. 8, pp. 1767–1780, 1972.

[15] R. Olshansky, "Mode-coupling effects in graded-index optical fibers," *App. Opt.*, vol. 14, no. 4, pp. 935–945, 1975.

[16] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, 1999.

[17] D. N.C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[18] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications—A key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[19] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Info. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[20] 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, 2011.

[21] 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, 2011.

[22] M. A. Jensen and J. W. Wallace, "A review of antennas and propagation for MIMO wireless communications," *IEEE Trans. Antennas Propag.*, vol. 52, no. 11, pp. 2810–2824, Nov. 2004.

[23] P. Almers, E. Bonek, A. Burr, C. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Özcelik, "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2007, p. 19070, 2007.

[24] K.-P. Ho and J. M. Kahn, "Frequency diversity in mode-division multiplexing systems," *J. Lightw. Technol.*, vol. 29, no. 24, pp. 3719–3726, 2011.

[25] A. Mecozzi, C. Antonelli, and M. Shtaif, "Coupled Manakov equations in multimode fibers with strongly coupled groups of modes," *Opt. Exp.*, vol. 20, no. 21, pp. 23436–23441, 2012.

[26] 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, 2013.

[27] S. Ö. 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, 2013.

[28] 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. 16680–16696, Aug. 2011.

[29] K.-P. Ho, "Exact model for mode-dependent gains and losses in multimode fiber," *J. Lightw. Technol.*, vol. 30, no. 23, pp. 3603–3609, Dec. 1, 2012.

[30] H. G. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[31] E. Schmidt, "Zur theorie de linearen und nichtlinearen integralgleichungen," *Math. Ann.*, vol. 63, pp. 433–476, 1907.

[32] W. Streifer, "Optical resonator modes—Rectangular reflectors of spherical curvature," *J. Opt. Soc. Amer.*, vol. 55, no. 7, pp. 868–877, 1965.

[33] D. A. B. Miller, "All linear optical devices are mode converters," *Opt. Exp.*, vol. 20, no. 21, pp. 23985–23993, 2012.

[34] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Boston, MA, USA: Now, 2004.

[35] D. Gesbert, M. Shafi, D.-S. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas in Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[36] G. L. Stüber, J. R. Barry, S. W. McLaughlin, Y. Li, M. A. Ingram, and T. G. Ptatt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.

[37] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.

[38] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, 2002.

[39] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ, USA: Prentice–Hall, 1995, Sec. 7.1.1.

[40] O. Alter, P. O. Brown, and D. Botstein, "Singular value decomposition for genome-wide expression data processing and modeling," *Proc. Nat. Acad. Sci.*, vol. 97, no. 18, pp. 10101–10106, 2000.

[41] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 931–940, Jul. 1998.

[42] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *J. SIAM Numer. Anal, Ser. B*, vol. 2, no. 2, pp. 205–224, 1965.

[43] A. Mecozzi and M. Shtaif, "The statistics of polarization-dependent loss in optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 14, no. 3, pp. 313–315, Mar. 2002.

[44] P. Lu, L. Chen, and X. Bao, "Statistical distribution of polarization dependent loss in the presence of polarization mode dispersion in single mode fibers," *IEEE Photon. Technol. Lett.*, vol. 13, no. 5, pp. 451–453, May 2001.

[45] A. Galtarossa and L. Palmieri, "The exact statistics of polarization-dependent loss in fiber-optic links," *IEEE Photon. Technol. Lett.*, vol. 15, no. 1, pp. 57–59, Jan. 2003.

[46] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization-mode dispersion in optical fibers," *Proc. Natl. Acad. Sci.*, vol. 97, no. 9, pp. 4541–4550, 2000.

[47] M. A. Nielsen, "Conditions for a class of entanglement transformations," *Phys. Rev. Lett.*, vol. 83, no. 2, pp. 436–439, 1999.

[48] A. G. Fox and T. Li, "Resonant modes in a maser interferometer," *Bell Sys. Tech. J.*, vol. 40, no. 2, pp. 453–488, 1961.

[49] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the Web," Stanford InfoLab, Stanford, CA, USA, Tech. Report, 1999.

[50] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Proc.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[51] M. Matthaiou, M. R. McKay, P. J. Smith, and J. A. Nossek, "On the condition number distribution of complex Wishart matrices," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1705–1717, Jun. 2010.

[52] R. W. Health and A. J. Paulraj, "Switching between diversity and multiplexing in MIMO systems," *IEEE Trans. Commun.*, vol. 53, no. 6, pp. 962–968, Jun. 2005.

[53] N. Bai, E. Ip, Y.-K. Huang, E. Mateo, F. Yaman, M.-J. Li, S. Bickham, S. Ten, J. Liñares, C. Montero, V. Moreno, X. Prieto, V. Tse, K. M. Chung, A. P. T. Lau, H.-Y. Tam, C. Lu, Y. Luo, G.-D. Peng, G. Li, and T. Wang, "Mode-division multiplexed transmission with inline few-mode fiber amplifier," *Opt. Exp.*, vol. 20, no. 3, pp. 2668–2680, 2012.

[54] Y. Jung, S. Alam, Z. Li, A. Dhar, D. Giles, I. P. Giles, J. K. Sahu, F. Poletti, L. Grüner-Nielsen, and D. J. Richardson, "First demonstration and detailed characterization of a multimode amplifier for space division multiplexed transmission systems," *Opt. Exp.*, vol. 19, no. 26, pp. B952–B957, 2011.

[55] S. Fan and J. M. Kahn, "Principal modes in multimode waveguides," *Opt. Lett.*, vol. 20, no. 2, pp. 135–137, 2006.

[56] M. B. Shemirani, W. Mao, R. A Panicker, and J. M. Kahn, "Principal modes in graded-index multimode fiber in presence of spatial- and polarization-mode coupling," *J. Lightw. Technol.*, vol. 27, no. 10, pp. 1248–1261, 2009.

[57] 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, 1991.

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

[59] A. J. Juarez, C. A. Bunge, S. Warm, and K. Petermann, "Perspectives of principal mode transmission in mode-division-multiplex operation," *Opt. Exp.*, vol. 20, no. 13, pp. 13810–13823, 2012.

[60] A. F. Garito, J. Wang, and R. Gao, "Effects of random perturbations in plastic optical fibers," *Science*, vol. 281, no. 5379, pp. 962–967, 1998.

[61] R. F. Shi, C. Koeppen, G. Jiang, J. Wang, and A. F. Garito, "Origin of high bandwidth performance of graded-index plastic fibers," *Appl. Phys. Lett.*, vol. 72, no. 25, pp. 3625–3627, 1999.

[62] S. Berdagué and P. Facq, "Mode division multiplexing in optical fibers," *Appl. Opt.*, vol. 21, no. 11, pp. 1950–1955, 1982.

[63] S. Murshid, B. Grossman, and P. Narakorn, "Spatial domain multiplexing: A new dimension in fiber optic multiplexing," *Opt. Laser Technol.*, vol. 40, pp. 1030–1036, 2008.

[64] C. P. Tsekrekos and A. M. Koonen, "Mode-selective spatial filtering for increasing robustness in mode group diversity multiplexing link," *Opt. Lett.*, vol. 32, no. 9, pp. 1041–1043, 2007.

[65] R. Ryf, S. Randel, A. H. Gnauck, C. Bolle, A. Sierra, S. Mumtaz, M. Esmaeelpour, E. C. Burrows, R.-J. Essiambre, P. J. Winzer, D. W. Peckham, A. H. McCurdy, and R. Lingle, "Mode-division multiplexing over 96 km of few-mode fiber using coherent 6×6 MIMO processing," *J. Lightw. Technol.*, vol. 30, no. 4, pp. 521–531, 2012.

[66] D. Marcuse, *Theory of Dielectric Optical Waveguide*, 2nd ed. New York, NY, USA: Academic, 1991.

[67] 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. 9484–9490, 2013.

[68] A. Galtarossa, L. Palmieri, M. Schiano, and T. Tambosso, "Measurement of birefringence correlation length in long, single-mode fibers," *Opt. Lett.*, vol. 26, no. 13, pp. 962–964, 2001.

[69] A. Galtarossa and L. Palmieri, "Spatially resolved PMD measurements," *J. Ligtw. Technol.*, vol. 22, no. 4, pp. 1103–1115, 2004.

[70] S. Kawakami and H. Tanji, "Evolution of power distribution in graded-index fibres," *Electron. Lett.*, vol. 19, no. 3, pp. 100–102, 1983.

[71] U. Levy, H. Kobrinsky, and A. A. Friesem, "Angular multiplexing for multichannel communication in a single fiber," *IEEE J. Quantum Electron.*, vol. QE-17, no. 11, pp. 2215–2224, Nov. 1981.

[72] S. Schöellmann, C. Xia, and W. Rosenkranz, "Experimental investigations of mode group diversity multiplexing on multimode fibre," presented at the Optical Fiber Communication Conf., Anaheim, CA, USA, 2006, paper OWR3.

[73] K. Kitayama, S. Seikai, and N. Uchida, "Impulse response prediction based on experimental mode coupling coefficient in a 10-km long graded-index fiber," *IEEE J. Quantum Electron.*, vol. QE-16, no. 3, pp. 356–362, Mar. 1980.

[74] N. Hanzawa, K. Saitoh, T. Sakamoto, T. Matsui, S. Tomota, and M. Koshiba, "Demonstration of mode-division multiplexing transmission over 10 km two-mode fiber with mode coupler," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2011, paper OWA4.

[75] S. D. Personick, "Time dispersion in dielectric waveguides," *Bell Sys. Tech. J.*, vol. 50, no. 3, pp. 843–859, 1971.

[76] M. L. Mehta, *Random Matrices*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2004.

[77] J. E. Cohen and C. M. Newman, "The stability of large random matrices and their products," *Ann. Probab.*, vol. 12, no. 2, pp. 283–310, 1984.

[78] M. B. Shemirani and J. M. Kahn, "Higher-order modal dispersion in graded-index multimode fiber," *J. Lightw. Technol.*, vol. 27, no. 23, pp. 5461–5468, 2009.

[79] M. Creutz, "On invariant integration over SU(N)," *J. Math. Phys.*, vol. 19, no. 10, pp. 2043–2046, 1978.

[80] P. A. Mello, "Averages on the unitary group and applications to the problem of disordered conductors," *J. Phys, A, Math. Gen.*, vol. 23, pp. 4061–4080, 1990.

[81] C. W. J. Beenaker, "Random-matrix theory of quantum transport," *Rev. Mod. Phys.*, vol. 69, no. 3, pp. 731–808, 1997.

[82] K.-P. Ho and W. Shieh, "Equalization-enhanced phase noise in mode-division multiplexed systems," *J. Lightw. Technol.*, vol. 31, no. 13, pp. 2237–2243, 2013.

[83] P. M. C. Andréief, "Note sur une relation entre les integrals définies des produits des functions," *Mém. de la Soc. des Sci. Phy. et Nat. de Bordeaux*, vol. II, no. 3, pp. 1–14, 1886.

[84] N. G. de Bruijn, "On some multiple integrals involving determinants," *J. Indian Math. Soc.*, vol. 19, pp. 133–151, 1955.

[85] 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.

[86] C. A. Tracy and H. Widom, "Level-spacing distributions and the Airy kernel," *Phys. Lett. B*, vol. 305, no. 1–2, pp. 115–118, 1993.

[87] C. A. Tracy and H. Widom, "Level-spacing distributions and the Airy kernel," *Commun. Math. Phys*, vol. 159, no. 1, pp. 151–174, 1994.

[88] F. Bornemann, "Asymptotic independence of the extreme eigenvalues of Gaussian unitary ensemble," *J. Math. Phys.*, vol. 51, p. 023514, 2010.

[89] E. Basor, Y. Chen, and L. Zhang, "PDEs satisfied by extreme eigenvalues distributions of GUE and LUE," *Random Matrices: Theory and Appl.*, vol. 1, no. 1, p. 1150003, 2011.

[90] C. Antonelli, A. Mecozzi, M. Shtaif, and P. J. Winzer, "Stokes-space analysis of modal dispersion in fibers with multiple mode transmission," *Opt. Exp.*, vol. 20, no. 11, pp. 11718–11733, 2012.

[91] F. Ferreira, D. Fonseca, A. Lobato, B. Inan, and H. Silva, "Reach improvement of mode division multiplexed systems with fiber splices," *IEEE Photon. Technol. Lett.*, vol. 25, no. 12, pp. 1091–1094, Jun. 2013.

[92] H. Bülow, H. Al-Hashimi, B. T. Abebe, and B. Schmauss, "Capacity and outage of multimode fiber with statistical bends," in *Proc. Opt. Fiber Commun. Conf.*, 2012, Paper OW3D2.

[93] L. Palmieri, L. Schenato, and A. Galtarossa, "The role of anisotropy in few-mode optical fibers," in *Proc. Opt. Fiber Commun. Conf.*, 2013, Paper OTu2G.3.
[94] 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, 1998.
[95] A. Galtarossa, L. Palmieri, and A. Pizzinat, "Optimized spinning design for low PMD fibers: An analytical approach," *J. Lightw. Technol.*, vol. 19, no. 10, pp. 1502–1512, 2001.
[96] A. Crisanti, G. Paladin, and A. Vulpiani, *Products of Random Matrices in Statistical Physics*. New York, NY, USA: Springer, 1993.
[97] M. Kang and M.-S. Alouini, "Capacity of MIMO Rician channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 112–122, Jan. 2006.
[98] C. M. Newman, "The distribution of Lyapunov exponents: Exact results for random matrices," *Commun. Math. Phys.*, vol. 103, pp. 121–126, 1986.
[99] A. Lobato, F. Ferreira, M. Kuschnerov, D. van den Borne, S. L. Jansen, A. Napoli, B. Spinnler, and B. Lankl, "Impact of mode coupling on the mode-dependent loss tolerance in few-mode fiber transmission," *Opt. Exp.*, vol. 20, no. 28, pp. 29776–29783, 2012.
[100] S. Warm and K. Petermann, "Splice loss requirements in multi-mode fiber mode-division-multiplex transmission links," *Opt. Exp.*, vol. 21, no. 1, pp. 519–532, 2013.
[101] A. Lobato, F. Ferreira, B. Inan, A. Adhikari, M. Kuschnerov, A. Napoli, B. Spinnler, and B. Lankl, "Maximum-likelihood detection in few-mode fiber transmission with mode-dependent loss," *IEEE Photon. Technol. Lett.*, vol. 25, no. 12, pp. 1095–1098, Jun. 2013.

Keang-Po Ho (S'91–M'95–SM'03) received the Ph.D. degree from the University of California, Berkeley, CA, USA, in 1995. He is currently with Silicon Image as a Principal Engineer and Senior Manager for the Baseband Algorithm Group. In Silicon Image (and SiBEAM before acquisition), he invented and implemented WirelessHD video area network using 60-GHz mm-wave antenna array, 3.8 Gb/s OFDM modem, and adaptive beam forming. He was the Chief Technology Officer and co-founder of StrataLight Communications (acquired by OpNext, Inc. for about US $170M). He has been with IBM T. J. Watson Research Center, Bellcore, the Chinese University of Hong Kong, and National Taiwan University. He has authored more than 200 journal and conference articles, 15 issued U.S. patents, several book chapters, and one book.

Joseph M. Kahn (M'90–SM'98–F'00) received the Ph.D. degree in physics from the University of California, Berkeley, CA, USA, in 1986. From 1987 to 1990, he was at AT&T Bell Laboratories, Crawford Hill Laboratory, in Holmdel, NJ, USA. He demonstrated multi-Gbit/s coherent optical fiber transmission systems, setting records for receiver sensitivity. From 1990 to 2003, he was on the faculty of the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. Since 2003, he has been a Professor of electrical engineering at Stanford University, Stanford, CA, USA. His research addresses fiber-based imaging, spatial multiplexing, rate-adaptive and spectrally efficient modulation and coding methods, and coherent detection and associated digital signal processing algorithms. Dr. Kahn received the National Science Foundation Presidential Young Investigator Award in 1991. In 2000, he cofounded StrataLight Communications, where he served as the Chief Scientist from 2000 to 2003.