


Effect of Higher-Order Modal Dispersion in Direct-Detection Mode-Division-Multiplexed Links

Anirudh Vijay , *Graduate Student Member, IEEE*, and Joseph M. Kahn, *Fellow, IEEE*

Abstract—Short-reach direct-detection links using mode-division multiplexing (MDM) in multimode fiber (MMF) can employ receiver-side optical signal processing (OSP), for example, by Mach-Zehnder meshes, to compensate for crosstalk among modes within mode groups, which have nearly equal propagation constants and couple strongly during propagation. In graded-index (GI) MMF, modes in the lowest-order groups, $\{LP_{01}\}$ and $\{LP_{11}\}$, also have nearly equal group delays. Modes in higher-order groups, such as $\{LP_{02}, LP_{21}\}$, do not, leading to modal dispersion (MD). To enable MDM in these higher-order mode groups, we propose to also employ transmitter-side OSP to launch signals into principal modes (PMs), which are free of MD to first order in frequency. We study the performance of the proposed scheme via simulation of realistic GI-MMFs based on experimental measurements by Carpenter et al. While PM transmission eliminates MD to first order in frequency, higher-order MD limits the ability of frequency-independent OSP to mitigate inter-symbol interference and crosstalk. We quantify the impact of higher-order MD, and show that either the PM coherence bandwidth or the group-delay spread within a mode group can predict the power penalty due to higher-order MD.

Index Terms—Direct detection, mode-division multiplexing, principal modes, Mach-Zehnder interferometer mesh, short-reach link, modal dispersion.

I. INTRODUCTION

MULTIPLEXING in spatial modes complements multiplexing in wavelength, time, phase, and polarization in scaling the capacity of fiber-optic communication systems. As information capacity scales linearly with the number of propagating modes, mode-division multiplexing (MDM) in multimode fibers (MMFs) is of fundamental interest and potential practical utility in future systems [1], [2], [3], [4], [5]. In MDM, as data signals on different modes couple during propagation, multi-input multi-output (MIMO) signal processing is required to recover the signals. Long-haul MDM systems typically employ coherent detection to recover the amplitude and phase of received signals and compensate for modal dispersion (MD) and mode coupling using digital signal processing (DSP). In

short-reach systems, MDM with direct-detection (DD) could potentially increase the capacity of installed MMFs. Typical DD systems do not recover signal phase, however, so mode coupling and MD must be compensated in the optical domain prior to photodetection.

Spatial and polarization modes in a graded-index (GI) MMF form groups with nearly equal propagation constants. In short-reach GI-MMF systems, intra-group mode coupling is strong and inter-group mode coupling is weak, making it sufficient to perform MIMO optical signal processing (OSP) in mode-group subspaces [6]. Structured OSP architectures employing Mach-Zehnder interferometer (MZI) meshes can realize any linear MIMO function, and can be programmed using a self-configuration algorithm, as shown by Miller [7]. Choutagunta et al. used MZI meshes at the receiver side to compensate for mode coupling in short-reach MDM systems [8], and described adaptive algorithms that can program the meshes faster than the self-configuration algorithm. The MZI meshes studied in [7], [8] are frequency-independent, so they cannot compensate for MD. While MIMO OSP devices with tunable frequency-dependent response (equivalently, with tunable temporal memory) could mitigate MD in principle, they are likely too complex for use in short-reach systems.

As an alternative method for mitigating MD, we propose to use frequency-independent MIMO OSP at the transmitter to launch data signals onto the principal modes (PMs) of the fiber, which are free of MD to first order in frequency. The PMs in MMF are a generalization of the principal states of polarization in a single-mode fiber (SMF), and form the least frequency-dependent modal basis set [9]. Fan and Kahn described the PMs in MMFs with mode coupling and dispersion [9], and explained how they differ from the eigenmodes of these MMFs. Several theoretical and experimental studies have since addressed the PMs [10], [11], [12], [13], [14]. Carpenter et al. performed extensive experimental characterization of the PMs in a variety of MMFs, characterizing key frequency- and time-domain properties, such as their coherence bandwidths and group delays [11]. Previous works have demonstrated mitigation of MD using frequency-independent OSP at the transmitter, implemented using spatial light modulators [15], [16], [17], but limited their scope to the transmission of a single data channel. We extend this approach to simultaneous transmission of multiple data channels and discuss efficient MIMO OSP using MZI meshes. This is the first time PM transmission has been proposed for the mitigation of MD in DD-MDM systems.

Manuscript received 27 May 2022; revised 31 October 2022; accepted 20 November 2022. Date of publication 5 December 2022; date of current version 16 March 2023. This work was supported by Maxim Integrated (now Analog Devices, Inc.). (Corresponding author: Joseph M. Kahn.)

The authors are with the E.L. Ginzton Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: avijay@stanford.edu; jmk@ee.stanford.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2022.3226704>.

Digital Object Identifier 10.1109/JLT.2022.3226704

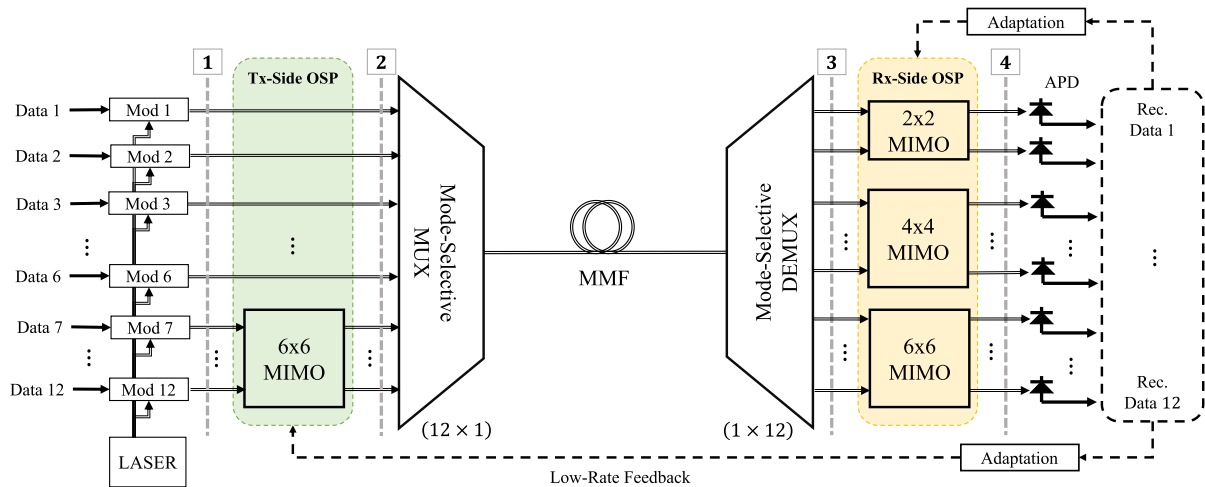


Fig. 1. Design of a short-reach DD-MDM link using transmitter- and receiver-side MIMO OSP. At the transmitter, a laser feeds multiple modulators. The modulated signals are optionally processed by MIMO OSP and multiplexed using a mode-selective multiplexer into an MMF. At the receiver end, a mode-selective demultiplexer separates the modes. The demultiplexed signals are processed mode-group-wise by MIMO OSP and their intensities are detected. The case of three mode groups with a total of 12 spatial and polarization modes is shown. Due to its inherent intra-group group-delay spread, only the third mode group, corresponding to data streams 7–12, requires transmitter-side MIMO OSP. Signal planes 1 through 4 are defined at the inputs and outputs of the OSP stages.

The proposed approach for mitigating MD by PM transmission will be limited by the finite coherence bandwidths of PMs imposed by higher-order MD, which is a generalization of higher-order polarization-mode dispersion. Shemirani et al. studied higher-order MD in GI-MMFs [13], presenting first- and second-order analytical transfer functions and an all-order numerical model to explain key phenomena, including pulse broadening, mode-pattern blurring, and a nonlinear input-output intensity response. As the length of an MMF link increases, the cumulative interplay of mode coupling and MD causes an enhancement of higher-order MD. In this work, we study the impact of higher-order MD on the proposed PM-based DD-MDM scheme via numerical modeling. We construct frequency-dependent matrix models to fit the extensive experimental measurements by Carpenter et al. [11], and study MDM performance at symbol rates of tens of Gbaud in fibers with lengths ranging from several hundred meters to several kilometers, representative of data centers and high-performance computing systems. This work is most relevant to multiplexing in modes belonging to groups that have nearly equal propagation constants but considerably unequal uncoupled group delays.

The key contributions of this paper are:

- 1) We propose an adaptive frequency-independent OSP scheme for PM transmission, i.e., launching signals onto PMs.
- 2) We study the performance of the scheme in short-reach systems, highlighting the impact of higher-order MD.
- 3) We show that either the PM coherence bandwidth or the group-delay spread within a mode group can predict the impact of higher-order MD in realistic GI-MMFs.

The remainder of this paper is organized as follows. Section II describes the proposed DD-MDM system architecture, reviews the frequency-dependent matrix model for MMF propagation, and describes transmission in PMs. Section III describes the proposed transmitter- and receiver-side adaptive

OSP for mitigating MD and mode coupling. Section IV presents system performance simulations, highlighting the impact of higher-order MD. Sections V and VI provide discussion and conclusions, respectively.

II. DIRECT-DETECTION MODE-DIVISION MULTIPLEXING IN THE PRESENCE OF MODAL DISPERSION

The proposed short-reach DD-MDM link architecture is shown in Fig. 1. It is similar to the design studied in [8], but with the addition of transmitter-side MIMO OSP. At the transmitter, a laser feeds multiple modulators to generate parallel intensity-modulated data signals. The modulated signals are subject to MIMO OSP and are multiplexed using a mode-selective multiplexer into the spatial and polarization modes of a short-reach MMF. At the receiver, the modes from the MMF are demultiplexed using a mode-selective demultiplexer, and the output signals are subject to MIMO OSP and detected by avalanche photodetectors (APDs).

The mode-selective multiplexer and demultiplexer can be realized using multi-plane light converters (MPLCs) or mode-selective photonic lanterns that map single-mode inputs to the spatial and polarization modes of the MMF [18], [19], [20], [21], [22]. Owing to the weakness of inter-group mode coupling in the MMF and the mode-selective multiplexer and demultiplexer, the MIMO OSP can be performed on each mode group separately. Consequently, the single laser source can be optionally replaced by multiple lasers, one for each mode group.

The DD-MDM scheme in [8] assumed that both the propagation constants and the group delays are the same for all the modes within a mode group, as in the $\{LP_{01}\}$ and $\{LP_{11a}, LP_{11b}\}$ mode groups with 2 and 4 modes, respectively, in GI-MMF. We extend that scheme to the scenario in which the propagation constants are nearly equal within a mode group but the group

delays are not, leading to MD. Higher-order mode groups typically exhibit this property. For concreteness, we focus on the third LP mode group $\{LP_{02}, LP_{21a}, LP_{21b}\}$ in GI-MMF with 6 spatial and polarization modes, corresponding to data channels 7 to 12 in Fig. 1. We address the MD in this third mode group by including transmitter-side MIMO OSP for this group in the system shown in Fig. 1.

A. System Propagation Model

Assuming negligible inter-group coupling, signal propagation in the $\{LP_{02}, LP_{21a}, LP_{21b}\}$ mode group can be modeled by $D \times D$ baseband-equivalent transfer matrices multiplying $D \times 1$ vectors of complex amplitudes, with $D = 6$. Assuming transmission at the zero-dispersion wavelength and short propagation distances, chromatic dispersion (CD) can be neglected. Mode-dependent loss (MDL) in the MMF is neglected, since short-reach links do not use optical amplifiers, and any splices or connectors are assumed to be well-aligned and few in number. The mode-selective multiplexer and demultiplexer are expected to be the most significant sources of MDL. To simplify the analysis, their MDL is neglected in Sections II and III. In Section IV, the impact of multiplexer and demultiplexer MDL on the proposed OSP scheme is studied.

Under these assumptions, propagation through MMF and the mode-selective multiplexer and demultiplexer, between planes 2 and 3 in Fig. 1, can be described in a multi-section model [23] by the product

$$\mathbf{M}_{2-3}(\omega) = \mathbf{U}_{K+1} \mathbf{\Lambda}_K(\omega) \mathbf{U}_K \cdots \mathbf{\Lambda}_2(\omega) \mathbf{U}_2 \mathbf{\Lambda}_1(\omega) \mathbf{U}_1, \quad (1)$$

where ω is the angular frequency, K is the number of fiber sections, the $\mathbf{\Lambda}_i(\omega) = \text{diag}\{e^{j(\omega-\omega_0)\tau_1}, e^{j(\omega-\omega_0)\tau_2}, \dots, e^{j(\omega-\omega_0)\tau_D}\}$, $i = 1, \dots, K$ are frequency-dependent diagonal unitary matrices modeling propagation in the sections, each with uncoupled modal group delays $\{\tau_1, \dots, \tau_D\}$, ω_0 is the center frequency, the \mathbf{U}_i , $i = 1, \dots, K+1$ are random frequency-independent unitary matrices modeling coupling between the sections, and $j = \sqrt{-1}$. The uncoupled group delays and the random unitary coupling matrices are chosen to model the fiber's physical properties, including refractive-index profile, bends and twists, and manufacturing processes such as spinning [24]. The random unitary matrices \mathbf{U}_1 and \mathbf{U}_{K+1} can model any phase shifts or intra-group crosstalk within the mode-selective multiplexer and demultiplexer, respectively.

Transmitter-side OSP, between planes 1 and 2 in Fig. 1, is modeled by a frequency-independent unitary matrix $\mathbf{M}_{1-2}(\omega) = \mathbf{U}$. Likewise, receiver side OSP, between planes 3 and 4, is modeled by a frequency-independent unitary matrix $\mathbf{M}_{3-4}(\omega) = \mathbf{V}^H$, where the superscript H represents Hermitian conjugate. The end-to-end transfer matrix between planes 1 and 4 is modeled by a product

$$\begin{aligned} \mathbf{M}_{1-4}(\omega) &= e^{-\alpha/2} \mathbf{M}_{3-4}(\omega) \mathbf{M}_{2-3}(\omega) \mathbf{M}_{1-2}(\omega) \\ &= e^{-\alpha/2} \mathbf{V}^H \mathbf{M}_{2-3}(\omega) \mathbf{U} \end{aligned} \quad (2)$$

where α is a loss parameter (in log power gain units) incorporating all mode-averaged losses in the system. Expression (2) captures the key linear effects of mode coupling and MD.

B. Transmission in Principal Modes

In the presence of mode coupling and MD, the best basis to represent signals in a multimode system is the set of PMs [6], which are free of MD to first order in frequency. The input PMs are the right eigenvectors of a group-delay operator $\mathcal{F}(\omega)$, which is related to the multiplexer-fiber-demultiplexer transfer matrix $\mathbf{M}_{2-3}(\omega)$ as [23]

$$\mathcal{F}(\omega) = -j \mathbf{M}_{2-3}^{-1}(\omega) \frac{\partial \mathbf{M}_{2-3}(\omega)}{\partial \omega}. \quad (3)$$

The eigenvalues of the input PMs are their group delays. Denoting the set of input PMs as the columns of a unitary matrix $\mathbf{U}^{(\text{PM})}(\omega)$, we can obtain a matrix of output PMs as $\mathbf{V}^{(\text{PM})}(\omega) = \mathbf{M}_{2-3}(\omega) \mathbf{U}^{(\text{PM})}(\omega)$. For a signal propagating between planes 2 and 3 in Fig. 1, the transfer function can be approximated to first order in frequency as

$$\mathbf{M}_{2-3}(\omega) \approx \mathbf{V}^{(\text{PM})}(\omega_0) \mathbf{\Lambda}(\omega) \mathbf{U}^{(\text{PM})H}(\omega_0), \quad (4)$$

where $\mathbf{\Lambda}(\omega) = \text{diag}\{e^{j(\omega-\omega_0)\tau_1^{(\text{PM})}}, e^{j(\omega-\omega_0)\tau_2^{(\text{PM})}}, \dots, e^{j(\omega-\omega_0)\tau_D^{(\text{PM})}}\}$, and $\{\tau_1^{(\text{PM})}, \dots, \tau_D^{(\text{PM})}\}$ are the group delays of the PMs at frequency ω_0 . In the approximation (4), only the middle diagonal matrix $\mathbf{\Lambda}(\omega)$ is frequency dependent and it captures first-order MD. The other two matrices in the product, namely, $\mathbf{U}^{(\text{PM})H}(\omega_0)$ and $\mathbf{V}^{(\text{PM})}(\omega_0)$, are frequency independent. They represent the mappings from the input to the input PMs in the basis of the ideal modes, and from the output PMs to the output in the basis of the ideal modes, respectively. Therefore, by performing a frequency-independent unitary transformation at the input with transfer matrix $\mathbf{M}_{1-2}(\omega) = \mathbf{U}^{(\text{PM})}(\omega_0)$, and a frequency-independent unitary transformation at the output with transfer matrix $\mathbf{M}_{3-4}(\omega) = \mathbf{V}^{(\text{PM})H}(\omega_0)$, we can diagonalize the system transfer matrix and obtain ideal transmission that is free of crosstalk and MD to first order in frequency. Using (4), we obtain an end-to-end transfer matrix

$$\mathbf{M}_{1-4}^{\text{ideal}}(\omega) = e^{-\alpha/2} \mathbf{V}^{(\text{PM})H}(\omega_0) \mathbf{M}_{2-3}(\omega) \mathbf{U}^{(\text{PM})}(\omega_0) \quad (5)$$

$$\approx e^{-\alpha/2} \mathbf{\Lambda}(\omega). \quad (6)$$

According to (6), in ideal PM transmission, the end-to-end system is a set of uncoupled parallel channels whose delays are given by the group delays of the PMs. The approximations in (4) and (6) are valid for input signals with bandwidths less

than the coherence bandwidth [10], [11] of the PMs.¹ In short-reach systems, the coherence bandwidths of lower-order PMs are of the order of tens to hundreds of GHz [11], sufficiently wide to support high-speed data transmission. Higher-order MD makes the PM coherence bandwidth finite, causing (4) and (6) to break down when the signal bandwidth exceeds the coherence bandwidth.

In the following two sections, we discuss how to implement adaptive PM transmission in the presence of MD using transmitter- and receiver-side MZI meshes, then we evaluate the performance of this scheme, highlighting limitations imposed by higher-order MD.

III. ADAPTIVE PRINCIPAL-MODE TRANSMISSION

The proposed adaptive transmission scheme employs a $D \times D$ MIMO OSP block after the modulators at the transmitter and a $D \times D$ MIMO OSP block before the photodetectors at the receiver. Ideally, the transmitter-side MIMO OSP block should map the modulated signals to the input PMs and the receiver-side MIMO OSP block should extract the received signals from the output PMs, effecting the transformations $\mathbf{U}^{(\text{PM})}(\omega_0)$ and $\mathbf{V}^{(\text{PM})H}(\omega_0)$, respectively. To achieve this, symbol-rate samples of received optical signals are taken, and an adaptive algorithm first tunes the transmitter-side MIMO OSP block and then tunes the receiver-side MIMO OSP block. The algorithm assumes that the receiver knows the transmitted sequences. In a practical system, this can be ensured by having dedicated training periods at regular intervals during data transmission. The required frequency of tuning depends on how fast the channel varies. The typical time scale for channel variation in short-reach MMF systems is a few milliseconds [25], [26]. The information needed to tune the transmitter-side OSP block is sent from the receiver to the transmitter via a low-rate feedback channel, as in [17]. The requirements for the low-rate feedback channel are discussed in Section V-C.

The MIMO OSP can be realized using various adaptive optical devices, such as MZI meshes [8], [27], multimode interference (MMI) couplers [28], [29], spatial light modulators [17], [21], [30], or sections of multicore fibers with phase shifters between them [31]. MZIs are an attractive option, as low-loss arrays can be fabricated in silicon photonics or other integration platforms. We use MZI meshes in the proposed scheme.

A. Transmitter-Side Mach-Zehnder Interferometer Mesh

The transmitter-side OSP block can be implemented using a triangular MZI mesh, as shown in Fig. 2. The mesh structure is similar to the receiver-side MIMO optical equalizer proposed

¹To determine the coherence bandwidth of a PM, the input mode pattern corresponding to the PM at the center frequency ω_0 is launched, and the resulting output mode pattern is recorded. Then, while keeping the input mode pattern fixed, the frequency ω is varied, the resulting output mode pattern at each ω is recorded, and its cross-correlation against the output mode pattern for frequency ω_0 is computed. We define the 1-dB coherence bandwidth of a PM as the range of ω (below and above ω_0) for which 10 times the base-10 logarithm of the magnitude of the cross-correlation is within 1 dB of the value obtained when $\omega = \omega_0$. In the numerical examples below, we will quantify the coherence bandwidth in units of $\omega/(2\pi)$ (Hz).

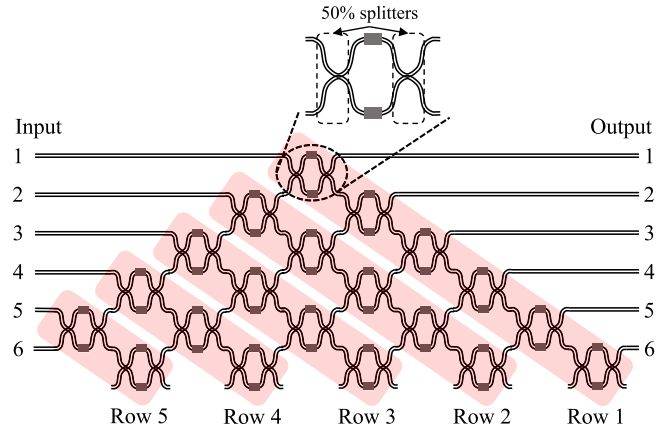


Fig. 2. Triangular MZI mesh for transmitter-side OSP. To realize a $D \times D$ MIMO block, $D - 1$ rows of MZIs are required. The i th row contains $D - i + 1$ MZIs. The example of $D = 6$ is shown.

in [8], and can realize any unitary transformation between the inputs and the outputs. To equalize the optical path lengths and losses, dummy MZI devices can be inserted appropriately [32]. The transfer matrix of the mesh is a deterministic function of the phase controls. An analytical derivation of the transfer matrix of an MZI mesh can be found in [7] and [8].

While the mesh can be theoretically set to produce the desired unitary transformation $\mathbf{U}^{(\text{PM})}(\omega_0)$, we require an adaptive algorithm that tunes the MZI phases to achieve that transformation. The working principle of the algorithm is derived directly from the concept of PMs. When a pulse is launched into an MMF in an arbitrary field pattern, it is subject to MD, and multiple copies arrive at the receiver with delays corresponding to the group delays of the PMs. If the pulse is instead launched into a PM, it arrives with a single delay. Therefore, the ideal configuration of the MZI mesh after tuning is one that ensures minimal pulse spreading or, equivalently, one that minimizes inter-symbol interference (ISI) among a sequence of pulses.

The algorithm must work in the presence of MD, mode coupling, and with random initial configurations of the transmitter- and receiver-side meshes. To meet these requirements, we use dispersion-sensitive training sequences and an ‘eye-opening’ metric proposed in [33]. The training sequences are transmitted on the data channels one by one and the corresponding row of the transmitter-side MZI mesh is tuned to increase the eye-opening metric, which is inversely related to ISI. Each successfully tuned row maps an input data stream onto an orthogonal input PM. The metric is calculated using the total optical intensity of all the modes in the mode group, or equivalently, the sum of all the APD photocurrents at the receiver. This physical quantity is used because it is independent of the configuration of the receiver-side mesh.

1) *Training Sequence and Eye-Opening Metric:* The training sequence consists of a string of L ‘off’ bits followed by a string of L ‘on’ bits as shown in Fig. 3. L must be greater than the memory length of the channel measured in units of the symbol interval T . To tune a row of the transmitter-side MZI mesh, the training sequence $I_{\text{in}}(t)$ is transmitted on the corresponding data channel.

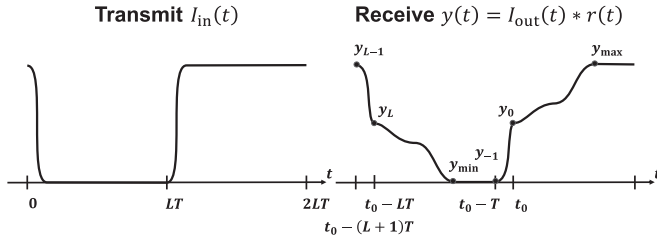


Fig. 3. Training sequence for tuning the transmitter-side OSP. The training sequence $I_{in}(t)$ has a duration of $2L$ symbols. The distorted received total optical intensity $I_{out}(t)$ is filtered by an impulse response $r(t)$ before being sampled. Six of those samples are used to estimate the ISI metric F .

The signal gets coupled into multiple PMs in the MMF and multiple copies arrive at the receiver with different delays. At the receiver, the total optical intensity $I_{out}(t)$ is obtained by summing all the photodetector outputs and filtering the sum by the receive filter. The filtered signal $y(t)$ is then sampled at the symbol rate to obtain the samples $y_l = y(t_0 + lT)$, $l = 0, \pm 1, \pm 2, \dots$, where t_0 is an initial sampling offset chosen by timing recovery in the receiver. The eye-opening metric is defined as [33]

$$F = (y_0 - y_{-1}) + (y_{L-1} - y_L) - (y_{\max} - y_{\min}). \quad (7)$$

The pairs of samples (y_{-1}, y_0) and (y_{L-1}, y_L) are those exhibiting the largest positive and negative excursions over an interval of length T , respectively, while y_{\min} and y_{\max} are the minimum and maximum samples, respectively. F is positive when the eye is open and negative when it is closed. Obtaining this metric as described is better than measuring an eye opening directly from an eye diagram, as this metric is usable even when the eye is closed. In the presence of noise, in order to obtain an accurate estimate of F , copies of the training sequence may be repeatedly transmitted and the resulting samples $\{y_l\}$ averaged. The receiver-side MIMO OSP is also a unitary transformation and, as a result, its setting does not affect $I_{out}(t)$. Consequently, the eye-opening metric is independent of the configuration of the receiver-side mesh.

2) *Tuning the Transmitter-Side Mach-Zehnder Interferometer Mesh*: Tuning the mesh involves setting the mesh transfer matrix \mathbf{U} to the matrix of input PMs $\mathbf{U}^{(PM)}(\omega_0)$ by performing a phase search for each controllable phase of the MZI devices. The method is detailed in Algorithm 1. The structure of the mesh is such that the i th row impacts columns $i, i + 1, \dots, D$ of the mesh transfer matrix \mathbf{U} . For example, the third row of the mesh in Fig. 2 affects columns 3 to 6 of \mathbf{U} but not columns 1 and 2. As a result of this property, the mesh can be tuned row by row, starting with the first row. Once the first row is tuned, it will have mapped the first data input to an input PM, thereby reducing the size of the matrix problem from $(D \times D)$ to $(D - 1 \times D - 1)$. The algorithm completes after running $D - 1$ row optimizations.

Each MZI device in the mesh is a 2×2 MIMO system with two phase controls, namely, the average phase θ_{avg} and the differential phase $\Delta\theta$. The differential phase affects the power splitting from the input to the output ports, and the average phase controls the common-mode phase of the signals at the output ports. The absolute phases of the column vectors of \mathbf{U} can be arbitrary. Therefore, the average phase controls of the first MZIs

Algorithm 1: Tuning the Transmitter-Side Mach-Zehnder Interferometer Mesh.

- 1: Random initialization of phase controls
 - 2: **for** row $i = 1, 2, \dots, D - 1$ **do**
 - 3: Activate i th input data stream
 - 4: Enumerate phase controls $\Phi = \{\Delta\theta^{(1)}, \theta_{\text{avg}}^{(2)}, \Delta\theta^{(2)}, \dots, \theta_{\text{avg}}^{(D-i)}, \Delta\theta^{(D-i)}, \theta_{\text{avg}}^{(D-i+1)}\}$
 - 5: $k \leftarrow 1$ {indexes Φ }
 - 6: **for** iteration $j = 1, 2, \dots, N_{\text{row,max}}$ **do**
 - 7: **while** $k \leq 2(D - i)$ **do**
 - 8: $k_1 = k$ {pick three phase controls}
 - 9: $k_2 = k + 1$ **if** $k + 1 \leq 2(D - i)$ **else**
 $k + 1 - 2(D - i)$
 - 10: $k_2 = k + 2$ **if** $k + 2 \leq 2(D - i)$ **else**
 $k + 2 - 2(D - i)$
 - 11: Tune phase controls $[\Phi(k_1), \Phi(k_2), \Phi(k_3)]$ in the search volume $\mathcal{S} \subseteq [0, \pi]^3$ sampled N_θ times per dimension {search volume \mathcal{S} is compressed after every pass}
 - 12: **for** $[\phi_1, \phi_2, \phi_3]$ in \mathcal{S} **do**
 - 13: Compute $F(\phi_1, \phi_2, \phi_3)$ {find the configuration of the three phases $\{\Phi(k_1), \Phi(k_2), \Phi(k_3)\}$ that maximize the F metric while the other phase controls are kept constant; Estimate of F is obtained by averaging over N_{avg} training sequences}
 - 14: **end for**
 - 15: $[\Phi(k_1), \Phi(k_2), \Phi(k_3)] \leftarrow \arg \max_{[\phi_1, \phi_2, \phi_3] \in \mathcal{S}} F$
 - 16: $k \leftarrow k + 3$
 - 17: **end while**
 - 18: $k \leftarrow k - 2(D - i)$ {wrap index for the next row iteration}
 - 19: **if** $|\Delta F| < \epsilon$ **then**
 - 20: $N_{\text{row}}^{(i)} \leftarrow j$
 - 21: **break** {exit if change in metric is below tolerance}
 - 22: **end if**
 - 23: **end for**
 - 24: **end for**
-

in each row, which control the absolute phases, need not be tuned. The last MZI in each row has only one output port, so its differential phase must be appropriately set to route all the optical power there. Hence, in the i th row of the mesh, there are $2(D - i)$ phase controls and the phase search is a $2(D - i)$ -dimensional optimization problem. Each phase can be in the range $[0, \pi]$. A search over the $2(D - i)$ -dimensional volume requires a single iteration per row but will be exponentially complex. On the other hand, an iterative one-dimensional search for each controllable phase, which will necessitate many communications from the receiver to the transmitter, will increase the latency of the algorithm. Owing to this trade-off between complexity of a single iteration of the algorithm and the communication latency in an iterative one-dimensional search, we choose to perform an iterative search for three phase controls at a time.

For each row in the mesh, Algorithm 1 enumerates the tunable phase controls and optimizes them three-at-a-time without overlap (except possibly the last set if $2(D - i)$ is not divisible by three, in which case the first few phase controls are included to complete the set). The search volume $\mathcal{S} \subseteq [0, \pi]^3$ contains N_θ values per dimension, corresponding to a total of N_θ^3 unique configurations. In the first iteration, the search volume is initialized to $\{0, \pi/N_\theta, 2\pi/N_\theta, \dots, (N_\theta - 1)\pi/N_\theta\}^3$, and the transmitter-side mesh is set to each of those search points one-at-a-time. For each configuration, at the receiver, the F metric is computed by averaging over N_{avg} training sequences. The transmitter-side mesh is set to the optimal configuration and the algorithm moves on to the next three-dimensional search in the row. Each pass over all $D - i + 1$ MZIs in the row involves $\lceil 2(D - i)/3 \rceil$ three-dimensional searches. After each pass, the search volume \mathcal{S} is compressed to a region around to the previous estimate. Each row optimization runs for a maximum of $N_{\text{row,max}}$ passes, or until the F metric changes by less than the tolerance ϵ , whichever happens first.

After all rows of the MZI mesh are tuned, the D input data streams will have been mapped to D orthogonal input PMs. The next step is to set the receiver-side MZI mesh to extract the data signals from the output PMs.

B. Receiver-Side Mach-Zehnder Interferometer Mesh

If the transmitter-side MZI mesh has been successfully set to $\mathbf{U}^{(\text{PM})}(\omega_0)$, then the receiver-side MZI mesh should ideally realize the frequency-dependent unitary transformation $\mathbf{V}^{(\text{PM})H}(\omega_0)$. This transformation can be realized using an MZI mesh adjusted using any of the tuning algorithms in [8]. In the proposed scheme, we have reserved a periodic training interval for tuning the transmitter-side mesh, so it is natural to reserve a subsequent interval for tuning the receiver-side mesh. We propose to use full-power training sequences, which were found to minimize receiver-side mesh tuning time in [8].

The row structure of the receiver-side MZI mesh is a mirror image of that of the transmitter-side MZI mesh in Fig. 2. The receiver-side MZI mesh is tuned row by row, starting with the first row. To tune a row, a full-power training sequence (simply a continuous-wave signal) is transmitted on the corresponding data channel. The objective of tuning is to route all the optical power to the corresponding output port of the mesh. Therefore, the received vector of detected intensities is scaled to unit norm and compared against a single-entry vector that is non-zero only at the index corresponding to the active data channel. The mean-squared error between the received vector (normalized to unit total power) and the single-entry vector is minimized by varying the phase controls in the MZI mesh. For a detailed description of the tuning algorithm, see Method 2 in [8].

IV. PERFORMANCE OF ADAPTIVE PRINCIPAL-MODE TRANSMISSION

In this section, we present simulations of the system shown in Fig. 1. The simulation parameters are provided in Table I. Signals are modulated by on-off keying (OOK) at a symbol rate $R_s = 56$ Gbaud. The center wavelength is $\lambda = 1310$ nm, close to the

TABLE I
SIMULATION PARAMETERS

Parameter	Symbol	Value
Number of modes (or) data channels*	D	6
Uncoupled differential group delay between LP_{02} and $LP_{21a,b}$ #	$\Delta\tau_0$	50 fs/m
Symbol rate	R_s	56 Gbaud
Modulation format		OOK
Received optical power per mode	P	-25 to -15 dBm
MMF link length	L_{MMF}	100 m to 5 km
Center wavelength	λ	1310 nm
Section length in multi-section model	ΔL	10 m
Correlation length in multi-section model	L_c	10 m
Maximum iterations for MZI row optimization	$N_{\text{row,max}}$	5
Number of test angles per dimension	N_θ	10
Averaging factor for F metric [†]	N_{avg}	100
Half-length of training sequence (in symbols)	L	2
APD responsivity	R	0.6 A/W
APD multiplication ratio	G	3.5
APD impact ionization coefficients ratio	k_a	0.18
Dark Current	I_d	10 nA
APD bandwidth	ΔF	48 GHz
TIA noise figure	F_n	1.5
TIA feedback resistor	R_F	100 Ω
Absolute temperature	T	300 K

* The simulations study the third mode group with 6 modes. A total of 12 modes are used in the MDM system. See Fig. 1.

This value is estimated using a numeric mode solver [34] assuming a parabolic refractive-index profile with 1% peak index contrast and random index perturbations of the order of 1.5×10^{-5} [35], [36].

[†] For tuning the transmitter-side MZI mesh only.

dispersion zero, so CD has only a minimal impact on system performance. We explicitly model only the 6×6 MIMO subsystem corresponding to the third mode group $\{LP_{02}, LP_{21a}, LP_{21b}\}$. We assume perfect timing recovery. At the receiver, the MZI mesh output is fed into an array of APDs, and the detected photocurrents are amplified by transimpedance amplifiers (TIAs). The APD parameters have been taken from [37]. The noise statistics describing the APD and the TIA are discussed in Appendix B.

In Section IV-A, we evaluate the performance of adaptive PM transmission in the presence of higher-order modal dispersion, assuming MDL is negligible. Then, in Section IV-B, we study the impact of intra-group MDL from the mode-selective multiplexer and demultiplexer on adaptive PM transmission.

A. Impact of Higher-Order Modal Dispersion

The simulations have two stages: a tuning stage and a data transmission stage. In the former, the MZI meshes are tuned using the adaptive algorithms described in Section II and in the latter, OOK data is transmitted on all 6 modes and the bit-error ratio (BER) is evaluated.

The fiber is modeled using a multi-section model [23] described by (1). The diagonal matrices $\mathbf{\Lambda}_i(\omega) = \text{diag}\{e^{j(\omega-\omega_0)\tau_1}, e^{j(\omega-\omega_0)\tau_2}, \dots, e^{j(\omega-\omega_0)\tau_D}\}$, $i = 1, \dots, K$, model modal dispersion, and the unitary matrices U_i ,

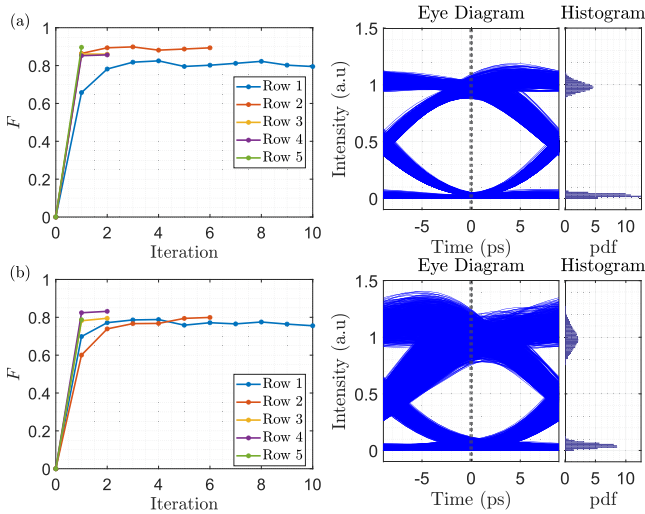


Fig. 4. Eye opening metric F as a function of iteration number during the tuning of the transmitter-side mesh, and eye diagram after tuning the transmitter- and receiver-side MZI meshes for (a) a GI-MMF of length 1350 m and a total received power of -18 dBm, and (b) a GI-MMF of length 3000 m and a total received power of -18 dBm.

$i = 1, \dots, K + 1$ model mode coupling. The uncoupled group delays in each section $\{\tau_1, \dots, \tau_D\}$ are deterministic and can be physically measured or numerically estimated given the refractive index profile of the GI-MMF. In the third mode group of a GI-MMF, MD is dominated by the group delay difference between the LP_{02} and $LP_{21a,b}$ spatial modes, and there is negligible delay difference between the polarization modes of each spatial mode. Hence, the vector of uncoupled group delays takes the form $\{\tau_1, \dots, \tau_D\} = \Delta\tau_0 \Delta L [0, 0, 1, 1, 1, 1]$, where ΔL is the length of a fiber section and $\Delta\tau_0$ is the differential uncoupled group delay per unit length between the LP_{02} and $LP_{21a,b}$ spatial modes.

The section length, the properties of the random coupling matrices \mathbf{U}_i , $i = 1, \dots, K$, and the differential uncoupled group delay $\Delta\tau_0$ are chosen to generate realistic models of commercially available OM-type GI-MMFs.² The generated models agree with the experimental results presented in [11]. Further discussion and details on the multi-section model can be found in Appendix A.

Using the multi-section model with the parameters in Table I, numerous fibers are generated with lengths ranging from 100 m to 5 km. For each fiber, the PMs, their group delays, and their coherence bandwidths are determined numerically using the group-delay operator (3). The transmitter- and receiver-side MZI meshes are tuned starting from random initial configurations. It is to be noted that the numerical computation of the PMs is not used in tuning the MZI meshes, but only in analyzing the results obtained.

Fig. 4 shows the optimization of the eye opening metric F using Algorithm 1 for two simulated fibers of lengths 1350 m and 3000 m for a received optical power of -18 dBm. The F metric

²We focus on higher-quality OM fibers, including OM3 and OM4, as their refractive-index profiles are closer to an ideal parabolic profile and they therefore exhibit increased intra-group coupling and negligible inter-group coupling.

is shown for the five row optimizations. Rows incorporating more MZI devices require more iterations to converge. It is observed that, owing to residual ISI from higher-order modal dispersion, the optimized F metric of the 3000 m-long fiber is lower than that of the 1350 m-long fiber. Also shown in Fig. 4 are the eye diagrams and the histogram of the sampled signal after tuning both the transmitter- and receiver-side meshes. The eye diagrams show the receiver noise and the residual ISI due to higher-order modal dispersion. As expected, the eye opening is smaller for the 3000 m-long fiber, and the corresponding histogram shows a significant increase in the standard deviation of the received ‘on’ bits for the same received optical power, indicating higher residual ISI due to higher-order modal dispersion. The asymmetry of the ISI in the ‘on’ and ‘off’ bits is due to a nonlinear intensity impulse response of higher-order modal dispersion, consistent with the observations in [13].

The statistics of the sampled signal and the eye opening can be used to estimate the BER. However, a more accurate estimate can be obtained using an enumeration method employing de Bruijn sequences of appropriate order (see Appendix B). The receiver sensitivity is estimated for a target BER of 2.4×10^{-4} by sweeping the received power between -25 dBm and -15 dBm per mode. Given the system parameters in Table I, the receiver sensitivity in the absence of MD is -21 dBm. The power penalty (in dB) due to MD is the difference between the receiver sensitivities (in dBm) with and without MD.

Fig. 5 shows the power penalty plotted against mean PM coherence bandwidth normalized to the symbol rate (Fig. 5(a)) and peak-to-peak group-delay spread normalized to the symbol period (Fig. 5(b)) for four different schemes using frequency-independent OSP:

- 1) Ideal PM transmission where the MIMO OSP block at the transmitter maps the input data streams onto the input PMs ($\mathbf{U} = \mathbf{U}^{(\text{PM})}(\omega_0)$) and the MIMO OSP at the receiver retrieves the signals from the output PMs ($\mathbf{V}^H = \mathbf{V}^{(\text{PM})H}(\omega_0)$), as explained in Section II-B.
- 2) Adaptive PM transmission as described in Section III. The transfer matrices of the MIMO OSP blocks in this scheme are expected to differ from those in ideal PM transmission because of noise and distortion of the training sequences by higher-order MD. This scenario represents practical implementation of PM transmission, and its performance is upper-bounded by that of the ideal PM transmission scheme.
- 3) Ideal optical equalization at the receiver to invert the channel transfer matrix at the center frequency ($\mathbf{V}^H = \mathbf{M}_{2-3}^{-1}(\omega_0)$). This scheme does not employ MIMO OSP at the transmitter, or equivalently, $\mathbf{U} = \mathbf{I}$.
- 4) Adaptive optical equalization at the receiver as described in [8] using full-power training sequences. Here too, $\mathbf{U} = \mathbf{I}$. The transfer matrix of this receiver-side MIMO OSP is expected to differ from that of the ideal optical equalization scheme due to noise. This scenario represents practical implementation of optical equalization at the receiver, and its performance is upper-bounded by that of the ideal optical equalization scheme.

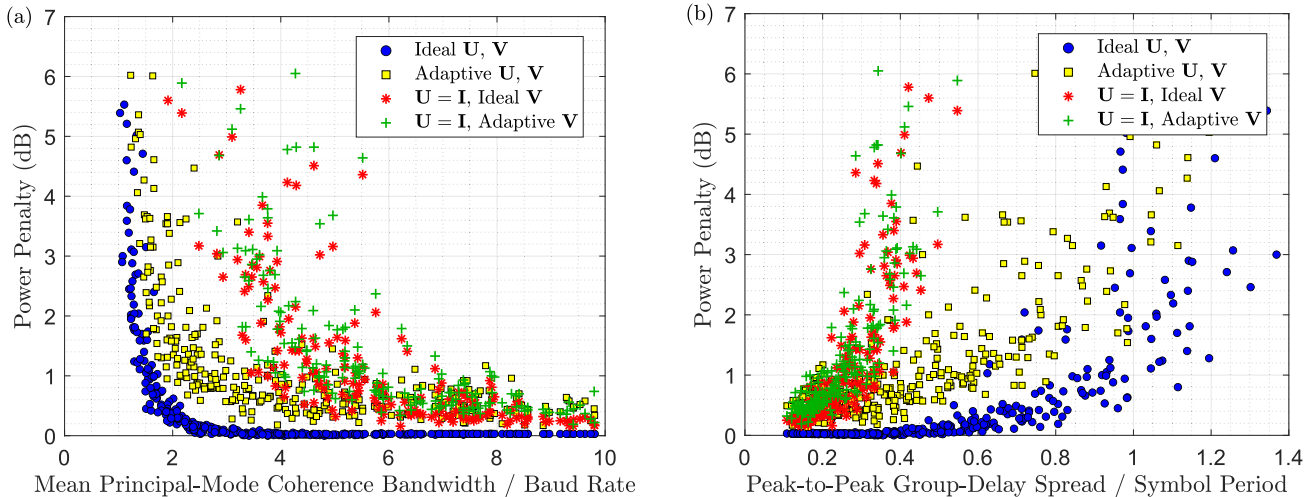


Fig. 5. Power penalty due to higher-order MD for various schemes plotted against (a) mean PM coherence bandwidth normalized to the symbol rate and (b) peak-to-peak group-delay spread normalized to the symbol period. Blue circles correspond to ideal PM transmission, $\mathbf{U} = \mathbf{U}^{(\text{PM})}(\omega_0)$, $\mathbf{V}^H = \mathbf{V}^{(\text{PM})H}(\omega_0)$. Yellow squares correspond to the adaptive PM transmission algorithm described in Section III. Red asterisks correspond to no transmitter-side OSP and ideal receiver-side optical equalization, $\mathbf{U} = \mathbf{I}$, $\mathbf{V}^H = \mathbf{M}_{2-3}^{-1}(\omega_0)$ [8]. Green pluses correspond to the adaptive receiver-side optical equalization algorithm using full-power training sequences described in [25].

While the ideal and adaptive receiver-based optical equalization schemes do not mitigate MD, they serve as baselines for comparison with the proposed adaptive PM transmission scheme.

The results in Fig. 5 can be understood as follows. Group-delay spread and mode coupling accumulate as signals propagate along the fiber. As the propagation length increases, apart from an increase in the peak-to-peak group-delay spread, the combination of mode coupling and unequal modal group delays causes the PMs to become frequency dependent, causing their coherence bandwidths to decrease. In the strong-coupling regime, the peak-to-peak and RMS group-delay spreads are proportional to the square-root of the propagation length, and the mean coherence bandwidth is inversely proportional to the square-root of the propagation length. In Fig. 5(a), shorter fibers (a few hundred meters long) fall on the right end of the plot and longer fibers (a few kilometers long) fall on the left end of the plot. On the other hand, in Fig. 5(b), shorter fibers fall on the left end of the plot and longer fibers fall on the right end.

When the PM coherence bandwidth is high, mode coupling is the only major impairment. In this regime, the power penalties of all four schemes are similar and are close to zero. As the coherence bandwidth decreases to about 4 times the symbol rate, the receiver-based optical equalization schemes (red asterisks and green pluses in Fig. 5), which do not address MD, are no longer sufficient and the power penalty increases rapidly due to ISI and crosstalk.

PM transmission (blue circles and yellow squares in Fig. 5(a)) removes first-order MD and the associated ISI and crosstalk, and hence performs better than receiver-based optical equalization. Any power penalty for these PM transmission schemes is dominated by the effects of higher-order MD. As the coherence bandwidth further reduces, residual ISI and frequency-dependent crosstalk due to higher-order MD increases, thereby

increasing the power penalty. From Fig. 5(a), it is clear that the limit of the technique is when the mean PM coherence bandwidth equals the symbol rate of the signals. The 2-dB penalty level occurs at 1.4 times the symbol rate for ideal PM transmission (blue circles) and at 1.8 times the symbol rate for adaptive PM transmission (yellow squares). The reach of PM transmission is 5-8 times that of the receiver-based optical equalization schemes. For the simulation parameters chosen in this study (Table I), the reach of ideal receiver-based optical equalization is about 550 m, while that of the adaptive and ideal PM transmission schemes are about 3000 m and 4500 m, respectively.

The performance of the adaptive PM transmission scheme is within 1 dB from that of ideal PM transmission for mean PM coherence bandwidths above 2.2 times the symbol rate. Aside from increasing signal distortion and crosstalk, higher-order MD also lessens the effectiveness of the tuning algorithms. The eye-opening metric for tuning the transmitter-side MZI mesh measures the ISI contributions from both first-order and higher-order MD. When the latter contributes significantly to the metric, the optimization no longer guarantees mapping of the transmitted signals to the input PMs. This explains the reduced performance of the adaptive PM transmission scheme compared to ideal PM transmission.

Fig. 5(b) plots the same data against the peak-to-peak group-delay spread. While the coherence bandwidth is a better predictor of the performance of PM transmission, the group-delay spread, which is typically easier to measure in a practical system, can also be used. In our simulations, we find an empirical relationship between the group-delay spread and the PM coherence bandwidth for links in the strong-coupling regime. When the propagation length exceeds 20 correlation lengths, i.e., $L_{\text{MMF}} > 20L_c$, we find that $\langle B \rangle \approx 1.36/\Delta\tau_{\text{pp}}$, where $\langle B \rangle$ is the mean coherence bandwidth and $\Delta\tau_{\text{pp}}$ is the peak-to-peak group-delay spread.

TABLE II
LINK BUDGET FOR 12 MODES IN THREE MODE GROUPS

Parameter	Value		Ref.
	Mode group 1, 2	Mode group 3	
Source power per mode	6.2 dBm	8.2 dBm	
Total source power*	18.1 dBm		
Transmitter coupling and routing loss	2 dB	2 dB	[38]
Modulator insertion loss	3 dB	3 dB	[39]
Modulation loss (10 dB extinction ratio)	2.6 dB	2.6 dB	
Transmitter-side MZI mesh loss	-	2 dB	[27]
Mode-selective MUX (MPLC) loss	4 dB	4 dB	[40]
MMF propagation loss at 0.6 dB/km	< 3 dB	< 3 dB	
Mode-selective DEMUX (MPLC) loss	4 dB	4 dB	[40]
Receiver-side MZI mesh loss	2 dB	2 dB	[27]
Receiver coupling and routing loss	2 dB	2 dB	[38]
Received optical power (per mode)	> -16.4 dBm		
Receiver sensitivity†	-20 dBm		
Unallocated power margin	> 3.6 dB		

* The system may use one laser or multiple lasers, one for each mode group.

† Referenced to the photodetector input, with 1 dB power margin.

Table II presents a sample link budget for the system in Fig. 1, which uses the 12 modes in the first three mode groups of the GI-MMF. Each signal is modulated at 56 Gbaud, providing an information bit rate of 50 Gbits/s assuming 12% overhead for forward error-correction coding. The total information bit rate is 600 Gbits/s. The link budget considers nominal losses for the various optical components. Assuming a total source power of 18.1 dBm, an unallocated power margin of at least 3.6 dB is obtained. As stated previously, the system can use one laser or multiple lasers, one for each mode group. The total optical power at the MMF input is 4.4 dBm, or 3.5 mW, which is well within a conservative eye-safety limit of 100 mW [41]. This suggests the possibility of further scaling the DD-MDM system capacity using wavelength-division multiplexing (WDM).

B. Impact of Multiplexer and Demultiplexer Mode-Dependent Loss

We have thus far neglected intra-group MDL in this paper. In the presence of intra-group MDL, the PMs are no longer orthogonal [12], and the unitary transformations of the MZI meshes cannot implement perfect PM transmission, causing additional power penalties. In this subsection, we consider intra-group MDL in the mode-selective multiplexer and demultiplexer and show that the adaptive PM transmission scheme using unitary transformations can tolerate values of MDL typically observed in these devices.

MPLC-based multiplexers and demultiplexers with low MDL have been reported [18], [19], [20]. The measurements in [19] indicate an MDL of 0.3 dB across the spatial modes of the

third mode group and a polarization-dependent loss of less than 0.2 dB. In our simulations, we assume a peak-to-peak intra-group MDL of 0.5 dB across the spatial and polarization modes of the third mode group in each of the multiplexer and demultiplexer.

The multi-section model in (1) can be modified to incorporate MDL:

$$\mathbf{M}_{2-3;\text{MDL}}(\omega) = \mathbf{V}_{\text{demux}}^H \boldsymbol{\Sigma} \mathbf{U}_{K+1} \boldsymbol{\Lambda}_K(\omega) \mathbf{U}_K \cdots \boldsymbol{\Lambda}_2(\omega) \mathbf{U}_2 \boldsymbol{\Lambda}_1(\omega) \mathbf{U}_1 \boldsymbol{\Sigma} \mathbf{U}_{\text{mux}}, \quad (8)$$

where \mathbf{U}_{mux} and $\mathbf{V}_{\text{demux}}^H$ are frequency-independent random unitary coupling matrices of the multiplexer and demultiplexer, respectively, $\boldsymbol{\Sigma} = \text{diag}\{e^{-\alpha_1/2}, e^{-\alpha_2/2}, \dots, e^{-\alpha_D/2}\}$ is a frequency-independent diagonal matrix modeling MDL, and $\alpha_1, \alpha_2, \dots, \alpha_D$ are the intra-group modal losses in log power gain units. We assume $\sum_{i=0}^D \alpha_i = 0$, as the mode-averaged loss is included in the factor $e^{-\alpha/2}$ in (2). We note that the model (8) can describe MDL arising in any components, such as connectors, at the transmitter and receiver.

The input PMs are the eigenvalues of the group-delay operator (3). With the model in (8), the input PMs at angular frequency ω_0 will now be the columns of $\mathbf{U}_{\text{MDL}}^{(\text{PM})}(\omega_0) = \mathbf{U}^{(\text{PM})}(\omega_0) \mathbf{U}_{\text{mux}}$. The output PMs will be the columns of $\mathbf{V}_{\text{MDL}}^{(\text{PM})}(\omega_0) = \mathbf{M}_{2-3;\text{MDL}}(\omega_0) \mathbf{U}_{\text{MDL}}^{(\text{PM})}(\omega_0)$. It is to be noted that $\mathbf{U}_{\text{MDL}}^{(\text{PM})}(\omega_0)$ is unitary and can be realized by the transmitter-side MZI mesh, but $\mathbf{V}_{\text{MDL}}^{(\text{PM})}(\omega_0)$ is non-unitary and cannot be realized by the receiver-side MZI mesh. In the case of weak MDL, however, a unitary transformation at the receiver can be close to optimal. Consider the singular value decomposition $\mathbf{V}_{\text{MDL}}^{(\text{PM})}(\omega_0) = \mathbf{V}_{\text{svd}}^H \boldsymbol{\Sigma}_{\text{svd}} \mathbf{U}_{\text{svd}}$, where \mathbf{U}_{svd} , $\mathbf{V}_{\text{svd}}^H$ are unitary matrices, and $\boldsymbol{\Sigma}_{\text{svd}}$ is a diagonal matrix of the singular values. The unitary matrix that best approximates $\mathbf{V}_{\text{MDL}}^{(\text{PM})}(\omega_0)$ is given by $\mathbf{V}_{\text{MDL}}^{(\text{PM})}(\omega_0) = \mathbf{V}_{\text{svd}}^H \mathbf{U}_{\text{svd}}$ [42], [43].

In the ideal PM transmission scheme, the transmitter-side mesh is set to the matrix of input PMs, $\mathbf{U} = \mathbf{U}_{\text{MDL}}^{(\text{PM})}(\omega_0)$, and the receiver-side mesh is set to the unitary matrix closest to the matrix of output PMs, $\mathbf{V}^H = \mathbf{V}_{\text{MDL}}^{(\text{PM})H}(\omega_0)$. In the adaptive PM transmission scheme, the transmitter- and receiver-side meshes are tuned using the scheme described in Section III. As the channel matrix is non-unitary, the receiver-side mesh tuning can be improved by employing several passes of the optimization algorithm.

Fig. 6 plots the power penalty against mean PM coherence bandwidth normalized to the symbol rate for no MDL and for multiplexer and demultiplexer peak-to-peak MDL values of 0.5 dB each, for both ideal and adaptive PM transmission. In adaptive PM transmission, the receiver mesh is tuned using five iterations of Method 2 described in [8]. With ideal PM transmission, MDL induces an additional penalty of 0.1 dB on average and 1.2 dB at maximum. With adaptive PM transmission, MDL induces an additional penalty of 0.1 dB on average and 3.4 dB at maximum. In spite of using only unitary OSP, PM transmission is able to effectively mitigate modal dispersion in the presence of MDL.

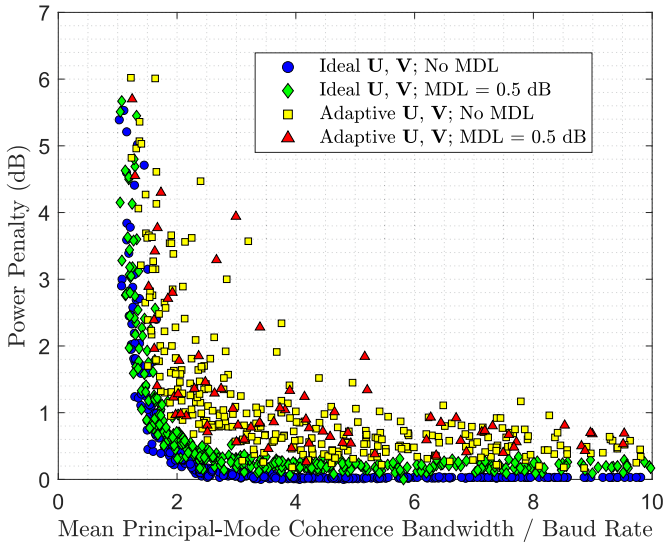


Fig. 6. Power penalty due to higher-order MD and MDL at the multiplexer and demultiplexer plotted against mean PM coherence bandwidth normalized to the symbol rate. Ideal and adaptive PM transmission with no MDL are represented by blue circles and yellow squares, respectively. Ideal and adaptive PM transmission with multiplexer and demultiplexer peak-to-peak MDL values of 0.5 dB each are represented by green diamonds and red triangles, respectively. The fiber realizations are different from those in Fig. 5.

V. DISCUSSION

In this section, we discuss some aspects of the proposed PM transmission scheme, namely, extension to WDM, training time, low-rate feedback channel, scaling to higher-order modes, and alternative OSP schemes.

A. Extension to Wavelength-Division Multiplexing

To extend the system to WDM, we replicate the transmitter and receiver subsystems in Fig. 1 for each wavelength. The wavelength multiplexer and demultiplexer should be multimode-compatible, and placed after the mode-selective multiplexers and before the mode-selective demultiplexers, respectively.

The link budget in Table II can be appropriately modified to accommodate WDM. Assuming additional losses of 1.7 dB each for the for WDM multiplexer and demultiplexer [44], and by reducing the unallocated power margin to 2 dB, the required source power increases to 19.9 dBm per wavelength (which can come from a single or multiple lasers, one for each mode group).

WDM operation with 4 wavelength channels placed $\pm 10, \pm 30$ nm from the zero-dispersion wavelength of 1310 nm is possible in principle. Such a system would support 48 data channels with a total throughput of 2.4 Tbits/s. The large channel spacing can obviate the need for precise temperature control.

Around 1310 nm, CD will be close to zero. Considering a high-estimate dispersion slope of 0.101 ps/(nm² km) [45], we obtain 4.8 ps of dispersion after 5 km of propagation for a 56 Gbaud channel 30 nm away from the zero-dispersion frequency. This is still much less than the spread due to MD and is not expected to cause substantial performance penalties.

We confirmed this through simulations, and obtained less than 0.2 dB additional penalty for ideal PM transmission.

B. Training Time

In Section IV, we presented the performance of the adaptive PM transmission scheme using MZI meshes. The results were obtained for tuning the meshes from random initial configurations to demonstrate the ability of the algorithm to work even in the presence of unmitigated MD and crosstalk. The training time for the transmitter-side MZI mesh can be estimated as

$$\begin{aligned}
 t_{\text{MZI-Tx}} = & 2LT \sum_{\text{row } i=1}^{D-1} \left[N_{\text{row}}^{(i)} \frac{2(D-i)}{3} \right] N_{\theta}^3 N_{\text{avg}} \\
 & + \Delta\tau_{\text{Rx-Tx}} \sum_{\text{row } i=1}^{D-1} \left[N_{\text{row}}^{(i)} \frac{2(D-i)}{3} \right] \\
 & + \Delta\tau_{\text{MZI-}\theta} \sum_{\text{row } i=1}^{D-1} \left[N_{\text{row}}^{(i)} \frac{2(D-i)}{3} \right] N_{\theta}^3, \quad (9)
 \end{aligned}$$

where $2LT$ is the period of the training sequence, N_{avg} is the number of times the F metric is averaged for each configuration, $N_{\text{row}}^{(i)} \leq N_{\text{row,max}}$ is the number of iterations of tuning the i th row, $\Delta\tau_{\text{Rx-Tx}}$ is the propagation time of the signal from the receiver to the transmitter through the feedback channel, and $\Delta\tau_{\text{MZI-}\theta}$ is the actuation time of the MZI phase shifters. The other parameters are defined in Section III-A and Table I. In (9), the first term corresponds to the duration of the training sequences, the second term corresponds to the communication latency between the receiver and the transmitter, and the third term corresponds to the time taken to change the phase configuration during tuning. In simulations, assuming an actuation speed of $\Delta\tau_{\text{MZI-}\theta} = 1$ ns [46] and a propagation time $\Delta\tau_{\text{Rx-Tx}} = 7.4$ μ s corresponding to $L = 1.5$ km [45], we estimate the tuning time of the transmitter-side mesh to be about 326 μ s. The tuning time for the receiver-side mesh is estimated to be about 20 μ s [8]. As described in Algorithm 1, we optimize the phase controls three-at-a-time. This is reflected in the N_{θ}^3 dependence of the first and last terms in (9). A three-dimensional search reduced the overall tuning time for the simulation parameters listed in Table I. For other sets of parameters, the tuning time of the algorithm may be minimized by using higher- or lower-dimensional searches.

Short-reach MMF channels change on timescales of several milliseconds [25], [26], [47], [48]. These changes can be tracked by tuning the MZI meshes periodically at intervals of several milliseconds. While the times required for initial tuning of the transmitter- and receiver-side meshes are estimated in the previous paragraph, successive tuning stages should take much less time, as they will start with better initial configurations.

C. Low-Rate Feedback Channel

To tune the transmitter-side MZI mesh, the adaptive algorithm for PM transmission requires communication between the receiver-side DSP circuitry and the transmitter. This feedback requires a bit rate of several Mbits/s. Bidirectional (duplex) links employing pairs of MMFs are widely deployed in local

area and data center networks [49]. These MMF pairs could provide forward and feedback channels for bidirectional PM transmission.

D. Scaling to Higher-Order Mode Groups

Throughout this paper, only the first three mode groups of the GI-MMF have been considered for data transmission. The proposed adaptive PM transmission scheme can be scaled to higher-order mode groups, which include progressively more modes. The number of MZI devices in an MZ mesh scales quadratically with the number of input ports. Annoni et al. discuss MIMO OSP transformations as large as 64×64 using MZI meshes [27]. The major challenges in scaling the MZI meshes are higher losses and power consumption, increased tuning time, and larger footprint.

While the first two mode groups are intrinsically free from MD, the third and higher-order mode groups exhibit MD, owing to unequal group delays between subgroups. Due to larger group delay differences between subgroups and higher sensitivity to perturbations causing mode coupling [25], the higher-order mode groups are expected to exhibit lower average PM coherence bandwidths for a given propagation length, making mitigation of MD more crucial.

E. Alternative Optical Signal Processing Schemes

The proposed MIMO OSP architecture using the transmitter- and receiver-side meshes perform frequency-independent OSP to address a frequency-dependent system. An alternative to this scheme would be to perform frequency-dependent OSP at the receiver [50]. While an all-order frequency-dependent MIMO equalizer might be difficult to realize, it is theoretically possible to perform first-order frequency-dependent OSP at the receiver using two MZI meshes and an array of tunable delay lines [7], [51], [52]. The matrix transfer function would take the form $\mathbf{V}^H \mathbf{\Lambda}(\omega) \mathbf{U}$, where \mathbf{V}^H and \mathbf{U} are the frequency-independent transfer matrices of the two meshes and $\mathbf{\Lambda}(\omega) = \text{diag}\{e^{j(\omega-\omega_0)\tau_1}, e^{j(\omega-\omega_0)\tau_2}, \dots, e^{j(\omega-\omega_0)\tau_D}\}$ is a diagonal matrix corresponding to the delays τ_1, \dots, τ_D of the delay lines. Alternatively, MMI couplers [29], [53] and delay lines could be used to realize the frequency-dependent MIMO OSP at the receiver.

VI. CONCLUSION

In this paper, we studied the impairments in short-reach DD-MDM systems caused by intra-group MD and mode coupling. We proposed a frequency-independent adaptive OSP scheme using MZ meshes to realize PM transmission, avoiding the effect of first-order MD. Through simulations, we evaluated the performance of the adaptive PM transmission scheme on the third mode group in GI-MMF with $D = 6$ modes, compared it with existing receiver-based optical equalization schemes, and studied the effect of higher-order MD on the performance. We showed that the mean PM coherence bandwidth and the group-delay spread can be used to predict the power penalty due to higher-order MD. Results in Fig. 5 show that ideal PM

transmission enables longer reach than receiver-based optical equalization schemes, the former requiring a mean PM coherence bandwidth of just 1.3 times the symbol rate as compared to more than 4 times the symbol rate for the latter. We also studied the performance of adaptive PM transmission in the presence of intra-group MDL from the mode-selective multiplexer and demultiplexer, showing that it causes only modest performance penalties. Finally, we discussed practical aspects of the adaptive algorithm including link budget, compatibility with WDM, and MZI mesh training time.

APPENDIX A

FIBER MULTI-SECTION MODEL

In this appendix, we discuss the details of the multi-section model (1) for simulating MMFs. Numerous works discuss modeling of multimode systems and their statistics [10], [23], [54], [55], [56], [57], [58], [59], [60]. [54], [55] describe power-coupling models of MMFs and therefore, cannot be directly applied to the study of PMs, which require field-coupling models. [10], [57] build models for mode coupling induced by bends and splices. [56], [58], [59] deal mainly with long-haul systems with tens to thousands of kilometers of propagation. In such systems, the fiber can be considered to be in the strong-coupling regime, and the RMS group-delay spread increases proportionally with the square-root of the length of the fiber. In this paper, we want a model that can capture all coupling regimes. In [60], a semi-analytic solution of the linear mode coupling equations is used to model propagation for the weak, intermediate and strong coupling regimes. While that model and those from [10], [57] are appropriate for our application in this work, we discuss a simpler way to extend the multi-section model in [23] for short-reach systems under the assumption that inter-group coupling and mode-dependent loss are negligible. One way of simulating such systems would be to use fewer sections (1 or 2) when the fiber is in the weak-coupling regime and more sections (10 s to 100 s) when the fiber is in the strong-coupling regime, with random full-unitary coupling matrices. Because the frequency order of the transfer matrix is proportional to the number of sections, however, we require a multi-section model that always uses many sections to accurately capture higher-order MD. To control the coupling strength between the sections, we must carefully generate the random coupling matrices $\{\mathbf{U}_i\}$ in (1).

To generate $D \times D$ random unitary matrices $\{\mathbf{U}_i\}$, one can first generate $D \times D$ random skew-Hermitian matrices $\{\mathbf{X}_{sh,i}\}$ and then exponentiate them: $\mathbf{U}_i = e^{\mathbf{X}_{sh,i}}$ (see Appendix A of [26]). The elements of $\mathbf{X}_{sh,i}$ can be sampled independently from a circularly symmetric, complex Gaussian distribution: $\mathbf{X}_{sh,i}[m, n] = -\mathbf{X}_{sh,i}^*[n, m] \sim \mathcal{CN}(0, \sigma_c^2)$; $m, n \in \{1, 2, \dots, D\}$. The standard deviation σ_c of the samples is a deterministic parameter and controls the strength of coupling. If σ_c is small, the entries in $\mathbf{X}_{sh,i}$ are more likely to be close to zero, thereby generating a random \mathbf{U}_i close to the identity matrix. This would correspond to weak coupling between fiber sections with most of the power from a mode in a section transferring to the same mode in the next section. The mode patterns at the end of consecutive sections would

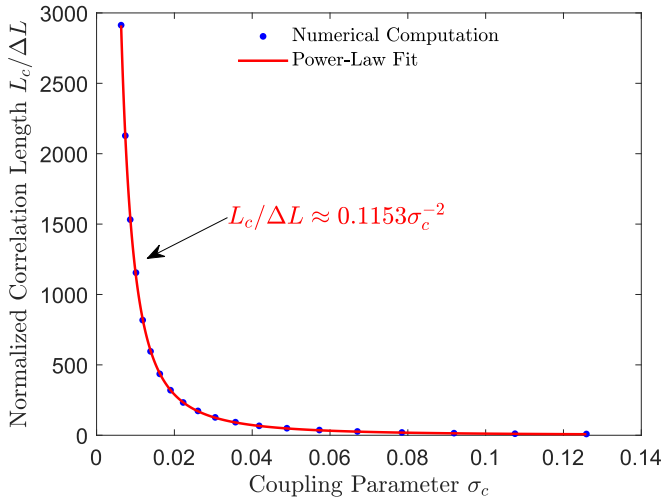


Fig. 7. Correlation parameters of the multi-section model for the 6×6 third mode group system. The figure shows the relationship between the correlation length L_c and the coupling parameter σ_c , which is the standard deviation of the elements in the skew-Hermitian-matrix exponent of the coupling matrices $\{\mathbf{U}_i\}$. The blue dots correspond to numerically generated data and the red curve corresponds to the best-fit power-law curve.

be highly correlated, and it would require propagation through many sections from the fiber input to make them uncorrelated from the input mode pattern. Similarly, if σ_c is large, then the coupling between sections would be strong and the mode pattern after just a few sections would be uncorrelated from the input mode pattern. To quantify this effect, we define a correlation metric

$$\gamma_c(k) = \mathbb{E} \left[\frac{1}{D} \text{tr} \{ \mathbf{U}_1^H \mathbf{U}_{1\dots k} \} \right], k = 1, 2, \dots \quad (10)$$

where $\mathbf{U}_{1\dots k} = \mathbf{U}_k \mathbf{U}_{k-1} \cdots \mathbf{U}_1$ is the cumulative product of the coupling matrices, and $\text{tr}\{\cdot\}$ is the matrix trace operation. We then define correlation length as the smallest propagation length for which the absolute value of γ_c falls below 0.5:

$$L_c = \Delta L \inf_{\substack{|\gamma_c(k)| < 0.5 \\ k \in \{1, 2, \dots\}}} k. \quad (11)$$

We numerically compute the relation between the correlation length L_c and the coupling parameter σ_c by generating numerous fibers with different σ_c and measuring L_c . The results are shown in Fig. 7. The best-fit curve gives us the approximate relationship $L_c \approx 0.1153 \Delta L / \sigma_c^2$. This inverse square law agrees with the observation that increasing the strength of the terms in the skew-Hermitian matrix makes the coupling stronger and decreases the propagation length needed to decorrelate the input mode pattern.

While the correlation length is deterministic, the generated fiber model is not. For a given fiber length, the PM group delays and coherence bandwidths are different for different realizations. Therefore, we generate many realizations and test the performance of the algorithm on each of them. We choose L_c , or equivalently σ_c , such that the numerically estimated PM coherence bandwidths fit the experimental data in [11]. In [11], the results are available for 100 m long MMFs only. While there are many combinations of σ_c and uncoupled group delay differences

$\Delta\tau$ that can result in similar coherence bandwidth values, our procedure ensures that the ensemble of fibers generated includes the real fibers in [11]. We plot the power penalty versus the normalized coherence bandwidth and normalized group-delay spread, instead of versus the link length, in order to further reduce the dependence of our results on our choice of simulation parameters.

The measurements in [11] were made at wavelengths near 1550 nm, whereas we apply them near 1310 nm. This is justified because the properties of the PMs to all orders in frequency depend on the uncoupled differential group delay $\Delta\tau_0$ and the mode-coupling coefficients, but do not depend on other properties, such as CD or MDL [13]. Numerical solution for the modal fields in the GI-MMF yields very similar values for the differential effective index and $\Delta\tau_0$ in the third mode group at the two wavelengths. While the PMs in the third mode group in a given fiber realization may not have identical properties at the two wavelengths, in an ensemble of fibers, the ensemble of these PMs should have very similar properties, especially when parameterized by the normalized bandwidth or normalized group-delay spread.

APPENDIX B

RECEIVER NOISE STATISTICS AND BIT-ERROR-RATIO COMPUTATION

In this appendix, we discuss the noise variances at the receiver and the method we employ to compute the BER performance. At the receiver, the noise in the photocurrents includes shot noise from the APD and the thermal noise from the TIA. The variances are given by

$$\sigma_{\text{th}}^2 = 4k_B T F_n \Delta f / R_F, \quad (12)$$

$$\sigma_{\text{shot}}^2(p) = 2qG^2 F_A (Rp + I_d) \Delta f, \text{ and} \quad (13)$$

$$\sigma_n^2(p) = \sigma_{\text{th}}^2 + \sigma_{\text{shot}}^2(p), \quad (14)$$

where σ_{th}^2 , $\sigma_{\text{shot}}^2(p)$, and $\sigma_n^2(p)$ are the thermal noise variance, shot noise variance and total noise variance, respectively. The shot noise variance is signal dependent, and is a function of the optical power p incident on the APD. k_B is the Boltzmann constant, T is the absolute temperature, F_n is the TIA noise figure, Δf is the filter bandwidth, R_F is the resistance of the TIA feedback resistor, q is the electronic charge, G is the APD gain, R is the APD responsivity, I_d is the dark current, and F_A is the excess noise factor. F_A is given by

$$F_A = k_a G + (1 - k_a) \left(2 - \frac{1}{D} \right), \quad (15)$$

where k_a is the ratio of the impact-ionization rate of holes to that of electrons in the APD. The relevant parameters are listed in Table I.

The signal-dependent shot noise also makes bit-error statistics signal dependent. Therefore, to accurately estimate the BER, we can use an enumeration method. In this method, the error probabilities of all possible input sequences are computed. The BER of the system is a weighted average of the individual error

probabilities with the weights equal to the probabilities of the sequences.

The transmitted symbols in each time slot and on each mode are independent and identically distributed. At the receiver, due to modal crosstalk and ISI due to MD, the received symbols are not independent of each other. The transmitter- and receiver-side OSP should eliminate most of the crosstalk and first-order MD. There will, however, be some residual crosstalk and ISI due to higher-order MD. We assume that the receiver does not perform any additional signal processing to mitigate these residual impairments and that it uses simple threshold detection to estimate the transmitted symbols. The threshold is common for all data channels and is decided by transmitting de Bruijn sequences [61]. A de Bruijn sequence of order n and an alphabet with size k is a cyclic sequence in which each possible subsequence of length n from the chosen alphabet occurs once and once only. By transmitting these sequences, we generate every possible pattern of ISI. Using de Bruijn sequences is an efficient way of implementing the enumeration method, as we can avoid transmitting all possible input sequences one after the other. For BER calculations, the alphabet size of the sequence is chosen based on the number of modes and the minimum order is based on the memory length of the channel. For $D = 6$ modes, we need an alphabet of size $2^6 = 64$. For the fibers that we generate, we have determined the memory length of the channel to be less than 40 ps. This corresponds to 2.24 symbols of ISI for a symbol rate of 56 GBaud. In our simulations, we set $n = 4$, which is sufficiently longer than the memory length of the channel. Once the de Bruijn sequences are generated with the 64-size alphabet, the symbols are converted to binary vectors of length $D = 6$. To estimate the BER, these vector sequences are transmitted and the received symbols are recorded. The noise statistics for these received signals are computed using (14). The optimum threshold $I_{th}^{(opt)}$ minimizes the BER for the given photocurrent and noise distributions:

$$I_{th}^{(opt)} = \arg \min_{I_{th}} \text{BER}(I_{th}), \quad (16)$$

and the BER is computed by averaging over each transmitted symbol as [62]

$$\text{BER}(I_{th}) = \frac{1}{MD} \times \sum_{i=1}^D \sum_{m=1}^M \frac{1}{2} \text{erfc} \left(\frac{Q_{i,m}(I_{th})}{\sqrt{2}} \right), \quad (17)$$

where M is the length of the de Bruijn sequence, $\text{erfc}(\cdot)$ is the complementary error function, and the Q factor of the symbol of the i th mode channel and m th position in the sequence is computed based on whether the transmitted symbol was a '0' or a '1':

$$Q_{i,m}(I_{th}) = \begin{cases} \frac{I_{th} - RGp_{i,m}}{\sigma_n(p_{i,m})}, & \text{if symbol } (i, m) \text{ is '0'} \\ \frac{RGp_{i,m} - I_{th}}{\sigma_n(p_{i,m})}, & \text{if symbol } (i, m) \text{ is '1'} \end{cases}. \quad (18)$$

ACKNOWLEDGMENT

The authors are grateful for helpful discussions with Karthik Choutagunta, Oleksiy Krutko, and Shaswat Mohanty.

REFERENCES

- [1] D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," *Nature Photon.*, vol. 7, no. 5, pp. 354–362, 2013.
- [2] G. Li, N. Bai, N. Zhao, and C. Xia, "Space-division multiplexing: The next frontier in optical communication," *Adv. Opt. Photon.*, vol. 6, no. 4, pp. 413–487, 2014.
- [3] R. Ryf et al., "Mode-division multiplexing over 96 km of few-mode fiber using coherent 6×6 MIMO processing," *J. Lightw. Technol.*, vol. 30, no. 4, pp. 521–531, Feb. 2012.
- [4] P. J. Winzer, "Making spatial multiplexing a reality," *Nature Photon.*, vol. 8, no. 5, pp. 345–348, 2014.
- [5] S. Ö. Arik, K.-P. Ho, and J. M. Kahn, "Optical network scaling: Roles of spectral and spatial aggregation," *Opt. Exp.*, vol. 22, no. 24, pp. 29868–29887, 2014.
- [6] K.-P. Ho and J. M. Kahn, "Mode coupling and its impact on spatially multiplexed systems," in *Optical Fiber Telecommunications, I*, VIT. Kaminow Li and A. Willner, Eds., vol. 17, New York, NY, USA: Elsevier, Amsterdam, The Netherlands, 2013, pp. 1386–1392.
- [7] D. A. B. Miller, "Self-configuring universal linear optical component," *Photon. Res.*, vol. 1, no. 1, pp. 1–15, 2013.
- [8] K. Choutagunta, I. Roberts, D. A. B. Miller, and J. M. Kahn, "Adapting Mach-Zehnder mesh equalizers in direct-detection mode-division-multiplexed links," *J. Lightw. Technol.*, vol. 38, no. 4, pp. 723–735, Feb. 2020.
- [9] S. Fan and J. M. Kahn, "Principal modes in multimode waveguides," *Opt. Lett.*, vol. 30, no. 2, pp. 135–137, 2005.
- [10] 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, May 2009.
- [11] J. Carpenter, B. J. Eggleton, and J. Schröder, "Observation of Eisenbud-Wigner-Smith states as principal modes in multimode fibre," *Nature Photon.*, vol. 9, no. 11, pp. 751–757, 2015.
- [12] A. A. 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. 13810–13824, 2012.
- [13] 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, Dec. 2009.
- [14] W. Xiong, P. Ambichl, Y. Bromberg, B. Redding, S. Rotter, and H. Cao, "Principal modes in multimode fibers: Exploring the crossover from weak to strong mode coupling," *Opt. Exp.*, vol. 25, no. 3, pp. 2709–2724, 2017.
- [15] E. Alon, V. Stojanovic, J. M. Kahn, S. Boyd, and M. Horowitz, "Equalization of modal dispersion in multimode fiber using spatial light modulators," in *Proc. IEEE Glob. Telecommun. Conf.*, 2004, pp. 1023–1029.
- [16] X. Shen, J. M. Kahn, and M. A. Horowitz, "Compensation for multimode fiber dispersion by adaptive optics," *Opt. Lett.*, vol. 30, no. 22, pp. 2985–2987, 2005.
- [17] R. A. Panicker, J. M. Kahn, and S. P. Boyd, "Compensation of multimode fiber dispersion using adaptive optics via convex optimization," *J. Lightw. Technol.*, vol. 26, no. 10, pp. 1295–1303, May 2008.
- [18] G. Labroille, P. Jian, N. Barré, B. Denolle, and J.-F. Morizur, "Mode selective 10-mode multiplexer based on multi-plane light conversion," in *Proc. Opt. Fiber Commun. Conf. Opt. Soc. Amer.*, 2016, Paper Th3E.5.
- [19] G. Labroille, B. Denolle, P. Jian, P. Genevieux, N. Treps, and J.-F. Morizur, "Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion," *Opt. Exp.*, vol. 22, no. 13, pp. 15599–15607, 2014.
- [20] N. K. Fontaine et al., "Packaged 45-mode multiplexers for a 50- μm graded index fiber," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.
- [21] N. K. Fontaine, R. Ryf, H. Chen, D. T. Neilson, K. Kim, and J. Carpenter, "Laguerre-Gaussian mode sorter," *Nature Commun.*, vol. 10, no. 1, pp. 1–7, 2019.
- [22] B. Huang et al., "All-fiber mode-group-selective photonic lantern using graded-index multimode fibers," *Opt. Exp.*, vol. 23, no. 1, pp. 224–234, 2015.
- [23] K.-P. Ho and J. M. Kahn, "Linear propagation effects in mode-division multiplexing systems," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 614–628, Feb. 2014.

- [24] L. Palmieri and A. Galtarossa, "Intramodal dispersion properties of step-index few-mode spun fibers," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 303–313, Jan. 2016.
- [25] K. Choutagunta and J. M. Kahn, "Dynamic channel modeling for mode-division multiplexing," *J. Lightw. Technol.*, vol. 35, no. 12, pp. 2451–2463, Jun. 2017.
- [26] K. Choutagunta, I. Roberts, and J. M. Kahn, "Efficient quantification and simulation of modal dynamics in multimode fiber links," *J. Lightw. Technol.*, vol. 37, no. 8, pp. 1813–1825, Apr. 2019.
- [27] A. Annoni et al., "Unscrambling light—automatically undoing strong mixing between modes," *Light Sci. Appl.*, vol. 6, no. 12, pp. e17110–e17110, 2017.
- [28] M. Bachmann, P. A. Besse, and H. Melchior, "General self-imaging properties in $n \times n$ multimode interference couplers including phase relations," *Appl. Opt.*, vol. 33, no. 18, pp. 3905–3911, 1994.
- [29] R. Tang, T. Tanemura, and Y. Nakano, "Integrated reconfigurable unitary optical mode converter using MMI couplers," *IEEE Photon. Technol. Lett.*, vol. 29, no. 12, pp. 971–974, Jun. 2017.
- [30] J. Carpenter, B. J. Eggleton, and J. Schröder, "110 × 110 optical mode transfer matrix inversion," *Opt. Exp.*, vol. 22, no. 1, pp. 96–101, 2014.
- [31] J. Zhou, J. Wu, and Q. Hu, "Tunable arbitrary unitary transformer based on multiple sections of multicore fibers with phase control," *Opt. Exp.*, vol. 26, no. 3, pp. 3020–3036, 2018.
- [32] D. A. B. Miller, "Self-aligning universal beam coupler," *Opt. Exp.*, vol. 21, no. 5, pp. 6360–6370, 2013.
- [33] R. A. Panicker and J. M. Kahn, "Algorithms for compensation of multimode fiber dispersion using adaptive optics," *J. Lightw. Technol.*, vol. 27, no. 24, pp. 5790–5799, Dec. 2009.
- [34] A. B. Fallahkhair, K. S. Li, and T. E. Murphy, "Vector finite difference modesolver for anisotropic dielectric waveguides," *J. Lightw. Technol.*, vol. 26, no. 11, pp. 1423–1431, Jun. 2008.
- [35] X. Chen, M. J. Li, and D. R. Powers, "Multimode optical fiber and systems comprising such fiber," U.S. Patent 8,971,683, Mar. 3, 2015.
- [36] K. Choutagunta and J. M. Kahn, "Designing high-performance multimode fibers using refractive index optimization," *J. Lightw. Technol.*, vol. 39, no. 1, pp. 233–242, Jan. 2021.
- [37] M. Huang et al., "56 GHz waveguide GE/SI avalanche photodiode," in *Proc. IEEE Opt. Fiber Commun. Conf. Expo.*, 2018, pp. 1–3.
- [38] L. Pavesi et al., "Silicon photonics III," *Topics Appl. Phys.*, vol. 122, pp. 1–36, 2016.
- [39] J. Witzens, "High-speed silicon photonics modulators," *Proc. IEEE*, vol. 106, no. 12, pp. 2158–2182, Dec. 2018.
- [40] S. Bade et al., "Fabrication and characterization of a mode-selective 45-mode spatial multiplexer based on multi-plane light conversion," in *Proc. IEEE Opt. Fiber Commun. Conf. Expo.*, 2018, pp. 1–3.
- [41] B. Buscaino, E. Chen, J. W. Stewart, T. Pham, and J. M. Kahn, "External vs. integrated light sources for intra-data center co-packaged optical interfaces," *J. Lightw. Technol.*, vol. 39, no. 7, pp. 1984–1996, Apr. 2021.
- [42] K. Fan and A. J. Hoffman, "Some metric inequalities in the space of matrices," *Proc. Amer. Math. Soc.*, vol. 6, no. 1, pp. 111–116, 1955.
- [43] J. B. Keller, "Closest unitary, orthogonal and hermitian operators to a given operator," *Math. Mag.*, vol. 48, no. 4, pp. 192–197, 1975.
- [44] H. Suhner and C. Cube, "Coarse wavelength division multiplexer-CWDM mux or demux, 4 channels, multi-mode," 2019. [Online]. Available: <http://literature.hubersuhner.com/Technologies/cubeoptics/specs/c-40mmrev-11-color-cube-cwdm-mux-or-demux-4-channel/>
- [45] "Corning ClearCurve OM2, OM3, and OM4 optical fiber product information," 2019. [Online]. Available: <https://www.corning.com/media/worldwide/coc/documents/Fiber/PI-1468-AEN.pdf>
- [46] N. Dupuis et al., "Nanosecond-scale Mach-Zehnder-based CMOS photonic switch fabrics," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 615–623, Feb. 2017.
- [47] X. Chen, J. He, A. Li, J. Ye, and W. Shieh, "Characterization and analysis of few-mode fiber channel dynamics," *IEEE Photon. Technol. Lett.*, vol. 25, no. 18, pp. 1819–1822, Sep. 2013.
- [48] X. Chen, J. He, A. Li, J. Ye, and W. Shieh, "Characterization of dynamic evolution of channel matrix in two-mode fibers," in *Proc. Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. Opt. Soc. Amer.*, 2013, Paper OM2C.3.
- [49] D. Coleman, "Optical trends in the data center," *ICT Today*, vol. 36, no. 5, pp. 16–22, 2015.
- [50] S. Ö. Arik and J. M. Kahn, "Direct-detection mode-division multiplexing in modal basis using phase retrieval," *Opt. Lett.*, vol. 41, no. 18, pp. 4265–4268, 2016.
- [51] A. Melloni et al., "Tunable delay lines in silicon photonics: Coupled resonators and photonic crystals, a comparison," *IEEE Photon. J.*, vol. 2, no. 2, pp. 181–194, Apr. 2010.
- [52] D. Pérez-López, E. Sánchez, and J. Capmany, "Programmable true time delay lines using integrated waveguide meshes," *J. Lightw. Technol.*, vol. 36, no. 19, pp. 4591–4601, Oct. 2018.
- [53] A. Rahim et al., "Finite impulse response filter using 4-port MMI couplers for residual dispersion compensation," *J. Lightw. Technol.*, vol. 30, no. 7, pp. 990–996, Apr. 2012.
- [54] R. Olshansky, "Mode coupling effects in graded-index optical fibers," *Appl. Opt.*, vol. 14, no. 4, pp. 935–945, 1975.
- [55] D. Marcuse, "Losses and impulse response of a parabolic index fiber with random bends," *Bell Syst. Tech. J.*, vol. 52, no. 8, pp. 1423–1437, 1973.
- [56] K.-P. Ho and J. M. Kahn, "Statistics of group delays in multimode fiber with strong mode coupling," *J. Lightw. Technol.*, vol. 29, no. 21, pp. 3119–3128, Nov. 2011.
- [57] A. A. Juarez, E. Krune, S. Warm, C. A. Bunge, and K. Petermann, "Modeling of mode coupling in multimode fibers with respect to bandwidth and loss," *J. Lightw. Technol.*, vol. 32, no. 8, pp. 1549–1558, Apr. 2014.
- [58] 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.
- [59] A. Mecozzi, C. Antonelli, and M. Shtaif, "Intensity impulse response of SDM links," *Opt. Exp.*, vol. 23, no. 5, pp. 5738–5743, 2015.
- [60] F. M. Ferreira, C. S. Costa, S. Sygletos, and A. D. Ellis, "Semi-analytical modelling of linear mode coupling in few-mode fibers," *J. Lightw. Technol.*, vol. 35, no. 18, pp. 4011–4022, Sep. 2017.
- [61] T. V. Aardenne-Ehrenfest and N. G. D. Bruijn, "Circuits and trees in oriented linear graphs," in *Classic Papers in Combinatorics*. Berlin, Germany: Springer, 2009, pp. 149–163.
- [62] P. M. Salehi and J. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2007.

Anirudh Vijay (Graduate Student Member, IEEE) received the B.Tech. and M.Tech. degrees in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2019. He is working toward the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA. His research interests include optical communications, mode-division multiplexing, and data-center applications.

Joseph M. Kahn (Fellow, IEEE) received the A.B., M.A., and Ph.D. degrees in physics from the University of California, Berkeley, Berkeley, CA, USA, in 1981, 1983, and 1986, respectively. During 1987–1990, he was with AT&T Bell Laboratories, Holmdel, NJ, USA. In 1989, he demonstrated the first successful synchronous (i.e., coherent) detection using semiconductor lasers, achieving record receiver sensitivity. In 1990–2003, he was on the Electrical Engineering and Computer Sciences Faculty with University of California. He demonstrated coherent detection of QPSK in 1992. In 1999, D. S. Shiu and Kahn published the first work on probabilistic shaping for optical communications. In the 1990s and early 2000s, Kahn and collaborators performed seminal work on indoor and outdoor free-space optical communications and multiinput multioutput wireless communications. In 2000, Kahn and K. P. Ho founded StrataLight Communications, whose 40 Gb/s-per-wavelength long-haul fiber transmission systems were deployed widely by AT&T, Deutsche Telekom, and other carriers. In 2002, Ho and Kahn applied to patent the first electronic compensation of fiber Kerr nonlinearity. StrataLight was acquired by Opnext in 2009. In 2003, he became a Professor of electrical engineering with the E. L. Ginzton Laboratory, Stanford University, Stanford, CA, USA. Kahn and collaborators have extensively studied rate-adaptive coding and modulation, and also digital signal processing for mitigating linear and nonlinear impairments in coherent systems. In 2008, E. Ip and Kahn (and G. Li independently) invented simplified digital backpropagation for compensating fiber Kerr nonlinearity and dispersion. Since 2004, Kahn and collaborators have been studying propagation, modal statistics, spatial multiplexing and imaging in multimode fibers, elucidating principal modes and demonstrating transmission beyond the traditional bandwidth-distance limit in 2005, deriving the statistics of coupled modal group delays and gains in 2011, and deriving resolution limits for imaging in 2013. His research interests include optical frequency comb generators, coherent data center links, rate-adaptive access networks, fiber Kerr nonlinearity mitigation, ultra-long-haul submarine links, and optimal free-space transmission through atmospheric turbulence. He was the recipient of the National Science Foundation Presidential Young Investigator Award in 1991. In 2000, he became a Fellow of the IEEE.