Adapting Mach–Zehnder Mesh Equalizers in Direct-Detection Mode-Division-Multiplexed Links

Karthik Choutagunta, *Student Member, IEEE*, Ian Roberts, *Student Member, OSA*, David A. B. Miller, *Fellow, IEEE, Fellow, OSA*, and Joseph M. Kahn, *Fellow, IEEE*

***Abstract*—A Mach–Zehnder mesh (MZM), which is comprised of a network of tunable 2×2 Mach–Zehnder interferometers and embedded photodetectors (PDs), can be used to perform arbitrary unitary matrix multiplications in the optical domain and compensate modal crosstalk in short-reach mode-division-multiplexed (MDM) links that use direct detection (DD). MZMs can be adapted using a self-configuration method, proposed by Miller, where multiple low-speed and low-power code sequences are superimposed on parallel high-speed information streams. We show that self-configuration in its original form is a sub-optimal equalization method for high-speed data transmission because adaptation based on detected code strengths is adversely impacted by low measurement signal-to-noise ratios and interference from the high-speed information streams. These impairments prevent the method from accurately tracking the millisecond-timescale modal dynamics of short-reach DD-MDM channels. We propose small modifications to the self-configuration method that can enable the MZM to track up to 10^8-fold faster channel dynamics. In particular, we show that replacing continuous equalization of low-power code sequences by periodic equalization of full-power training signals and using special optimization methods can yield faster MZM tuning. We also discuss the tradeoffs between MZM architectures that embed PDs inside the mesh and those that have PDs at the output ports only. Our results indicate that optimally designed MZMs and their associated control methods can increase the information capacity of short-reach multimode optical fiber links.**

***Index Terms*—Adaptive equalization, direct detection, Mach-Zehnder mesh, mode-division multiplexing, short-reach links.**

I. Introduction

MODE-DIVISION multiplexing (MDM) in optical fibers can enable the information capacity of links to scale with the number of propagating modes. Most of the recent MDM research has been focused on systems using coherent detection for long-haul communications (100s to 1000s of km), where both the amplitude and phase of transmitted signals are recovered at the receiver using the aid of a local oscillator and equalization is performed in the digital domain [2]–[4]. In this paper, we study short-reach MDM systems (100 m to 2 km) using direct detection (DD) that can be potentially deployed in data-center settings. In DD systems, the detected photocurrents are proportional to the total received optical powers because the received signals are not mixed with a local oscillator. Since the phase and frequency information of the transmitted signals is lost, DD-MDM systems detect only modal intensities and therefore must equalize signals using optical multiple-input multiple-output (MIMO) processing prior to photodetection. This paper is about the unique algorithmic and practical challenges associated with DD-MDM.

DD systems using multi-mode propagation have traditionally relied on mode-group-division multiplexing (MGDM) where the mode groups in graded-index multimode fiber (MMF) form parallel subchannels of communication [5]–[8]. In these systems, all of the modes belonging to a single mode group convey a common information stream. MGDM is a suboptimal approach to spatial multiplexing because the information capacity scales approximately as $\sqrt{D}$, where D is the number of spatial and polarization modes supported by the MMF.

An alternative approach that can that can recover all D degrees of freedom in DD-MDM systems was proposed by Arik *et al.* [9]. In particular, phase retrieval (implemented by an algorithm based on convex optimization) can be used to estimate a complex-valued MIMO channel by transmitting known training vectors and detecting the intensities of received modes. The MIMO channel estimate can be mathematically inverted and physically realized by an optical equalizer to undo modal crosstalk. While this work showed that it was theoretically possible to use all spatial and polarization modes to communicate over a directly detected MIMO channel, the proposed equalization methods have the drawbacks of high computational complexity and the requirement for high-speed analog-to-digital converters. Some progress has been made on reducing the complexity of phase retrieval-based DD-MDM [10], but simpler methods for adapting optical MIMO equalizers are still desired.

A mesh of Mach-Zehnder interferometers (MZIs) can realize arbitrary unitary transformations and act as an optical MIMO equalizer [1]. The phase shifts in MZ meshes (MZMs) can be configured with low complexity using a self-configuration method, which is discussed in more detail in Section III. Preliminary experiments have demonstrated the concept [11], but the best method for configuring MZMs for high-speed data transmission is still an open challenge. Short-reach MDM optical links experience millsecond-timescale modal dynamics

Manuscript received August 27, 2019; revised November 1, 2019; accepted November 4, 2019. Date of publication November 7, 2019; date of current version February 12, 2020. This work was supported by Maxim Integrated, a Stanford Graduate Fellowship, and by the Air Force Office of Scientific Research under award number FA9550-17-1-0002. *(Corresponding author: Karthik Choutagunta.)*

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

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

Digital Object Identifier 10.1109/JLT.2019.2952060

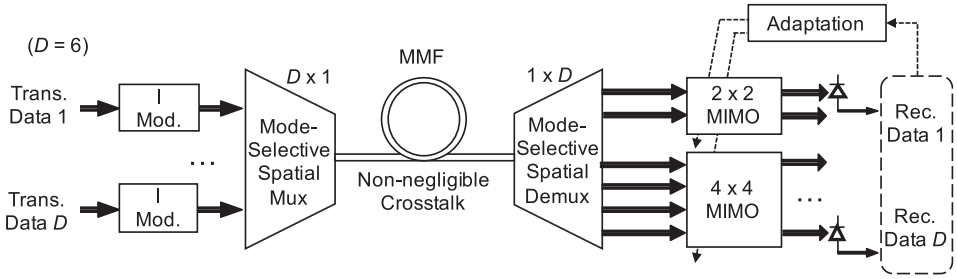

Fig. 1. Canonical architecture for a short-reach mode-division-multiplexed optical link using direct detection. D parallel information streams are independently modulated and multiplexed onto a multimode fiber using a mode-selective spatial multiplexer. At the receiver, the modes are demultiplexed by another mode-selective spatial demultiplexer and are processed by adaptive optical equalizers in the mode-group subspaces. The case of $D = 6$ parallel information streams is shown.

[12]–[14], so MZMs must converge and track evolving channels on the sub-millisecond timescale while maintaining low hardware and algorithmic complexity to be practically feasible.

In this paper, we present two main contributions:

1) We rigorously analyze the self-configuration method for adapting MZMs in DD-MDM systems by deriving expressions for the interference and noise impairments and simulate the performance of this method for an example link.
2) We propose optimized variants of the self-configuration method and show that these modified methods are capable of tracking faster channel dynamics.

This paper is organized as follows. Section II reviews the canonical link architecture for DD-MDM systems and the goals of optical MIMO processing using MZMs. Section III discusses the self-configuration method to adapt MZMs and simulates its performance for an example DD-MDM link using $D = 4$ modes. Section IV proposes several variants of the self-configuration method and discusses their tradeoffs. Finally, Sections V and VI provide discussions and conclusions, respectively.

II. Direct-Detection Mode-Division-Multiplexed Links

A. DD-MDM Link Architecture

The canonical architecture of short-reach DD-MDM links is illustrated in Fig. 1. The transmitter splits light from a single laser source into D different paths and modulates their intensities by independent data streams. The modulated signals are coupled into a MMF using a mode-selective spatial multiplexer. The mode-selective spatial multiplexer can be implemented as a multiplane light converter (MPLC) that maps spatially separated single-mode fiber inputs into spatially overlapping, but orthogonal, modes of an MMF [15]–[17]. The MMF is assumed to be short such that chromatic and modal dispersions can be neglected. Graded-index MMFs are preferred to step-index MMFs because the former have less modal dispersion within mode groups than the latter. Furthermore, random perturbations such as bends and twists along the fiber typically cause strong random unitary mode coupling within mode groups but weak mode coupling across different mode groups. When the modes are demultipexed at the receiver (e.g., by an MPLC operating in the reverse direction), the modal crosstalk can be processed independently in the mode-group subspaces. Since the group delays of the modes within each mode group can be assumed to be equal after propagation through the short length of fiber, the impulse response of the channel is short. Therefore, the optical MIMO equalizers for each mode group do not need to have temporal memory and can be implemented as single-tap matrix multiplications.

Fig. 1 shows an exemplary MMF link supporting $D = 6$ spatial and polarization modes, where the crosstalk in the $N = 2$ polarization modes of the LP_{01} mode group is processed by a 2×2 optical MIMO equalizer, and the crosstalk in the $N = 4$ spatial and polarization modes of the LP_{11} mode group is processed by a separate 4×4 optical MIMO equalizer.[1] Alternatively, the crosstalk in the entire set of 6 modes can be processed by a single 6×6 optical MIMO equalizer, but processing the modes in the mode-group subspaces is more computationally efficient because we assume negligible scattering from one mode group to another.

B. Mach-Zehnder Mesh as an Optical MIMO Equalizer

Optical MIMO equalizers can be physically realized using adaptive optics formed from cascades of multimode interference (MMI) coupler arrays [18]–[20], MZMs [1], [11], or from sections of multicore fibers and phase shifters [21]. MZMs are an attractive choice for implementing optical MIMO functions because they can be constructed as a relatively simple network of MZI coupler building blocks. They can be fabricated easily in integrated silicon photonics and scaled to large sizes. MZMs can either be arranged either in a triangular [22] or a rectangular configuration [23]. While rectangular MZMs have a more uniform optical path length from input ports to output ports and are more robust to optical losses, they cannot be easily controlled using the self-configuration method of Section III. It is possible to equalize optical path lengths and path losses in triangular MZMs by inserting dummy MZIs [24]. Cascaded binary tree configurations offer an alternative self-configuring architecture [24], though these have not yet been investigated extensively.

In this paper, we will focus on a particular mode group with N modes and consider MZMs with N input and output ports, as

[1]We use D to refer to the total number of polarization and spatial modes supported by an MMF link across all mode groups, and use N to refer to the size of a particular mode group.

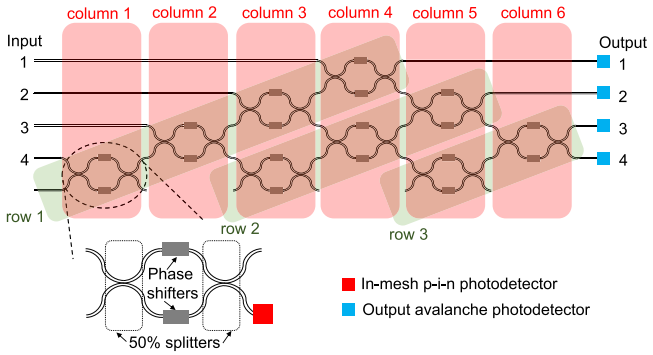

Fig. 2. A triangular Mach-Zehnder Mesh (MZM) with $N = 4$ input and output ports. The rows of the MZM are colored in green whereas the columns are in pink. The inset shows the structure of an individual MZI coupler, which may contain a low-speed in-mesh p-i-n photodetector (PD) for adapting the MZM. The output ports contain high-speed avalanche PDs with large gain-bandwidth product because they lie on the data signal paths.

shown in Fig. 2. Each mode group is independently processed by a separate MZM, as shown in Fig. 1. Furthermore, we assume each MZM is implemented as in a triangular configuration because their phase shifts can be easily controlled using the self-configuration method. There are $(N^2 + N - 2)/2$ MZIs per MZM.[2] For notational convenience, we define a *row* of the MZM as a set of MZIs which are chained from left to right, i.e., the top right output port of an MZI is connected to the bottom left input of the next MZI. We define a *column* of the MZM by selecting MZs that line up vertically in the triangular arrangement. The MZIs belonging to the same column are not directly connected to each other.

We define the notation MZI (r, c) to refer to the MZI in row r and column c of the MZM. This notation will be useful when describing MZM adaptation techniques in Sections III and IV. Each MZI coupler (r, c) has two programmable controls, the common phase $\theta_{av}^{(r,c)}$ and the differential phase $\Delta\theta^{(r,c)}$. These phase shifts can be programmed to optically implement any $N \times N$ unitary transmission matrix. The details of the transmission matrix of the MZI couplers are discussed in Appendix VI. Note that various alternate forms are possible for such devices with phase shifters in different arms. In all cases, we need to have at least one phase shifter on a MZI arm to control the "split ratio" of the coupler, and one other phase shifter that allows control of another phase difference.

As shown in Fig. 2, we consider two types of triangular MZMs in this paper: those with embedded in-mesh photodetectors (PDs), and those with PDs at the output ports only. The in-mesh PDs are located at the bottom-right output port of each MZI coupler (r, c). These PDs are mostly transparent and tap a small fraction γ of the incident optical powers, which is used in adaptive algorithms to tune the MZM to optically implement a desired unitary matrix. The in-mesh PDs can be contactless integrated photonic probes [25] or p-i-n diodes. For reasons discussed later in Section III, the in-mesh PDs can have low-speed transimpedance amplifiers (TIAs) with low noise figures. On the other hand, the PDs at the output ports of the mesh are avalanche PDs (APDs), which are higher-speed because they lie on the data path.

[2]The MZIs in the bottom of Fig. 2 can be replaced by simple phase shifters.

III. Evaluation of the Self-Configuration Method

The goal of the adaptive methods presented in this paper is to set all the MZI coupler phase shifts of a triangular MZM to separate N parallel data signals that have been randomly mixed by propagation through a fiber channel. Random mode coupling will cause each data signal to appear distributed over the MZM input ports. When the MZM is properly configured, each data signal will appear at a separate output port with minimal crosstalk from the remaining data signals. In this section, we consider the self-configuration method, proposed by Miller in [1], and rigorously quantify its performance in a DD-MDM system.

A. Review of Self-Configuration Method

We first briefly review the self-configuration concept before discussing its details more rigorously in Section III-B. Each of the N information streams is modulated by slowly varying pilot tones (or codes) to uniquely identify each data signal [1]. The self-configuration method adapts the MZM by detecting the strength of the codes. When the MZM is fully adapted and successfully separates the codes (i.e., forms a one-to-one mapping between code i and output port i), then the underlying information streams will also be separated (i.e., information stream i will only appear at output port i).

The MZM contains various mostly transparent PDs that are embedded in the mesh at strategic locations. These PDs are used to tap a portion of impinging power to determine the strength of codes present. Since the codes are slowly varying, the in-mesh PDs can have low bandwidth. The self-configuration method proceeds by sequentially optimizing the phase shifter angles of the MZs in the first row (see Fig. 2) so as to maximize the detected power of the first code at the first mesh output (thereby also maximizing the power of the first information stream at the first output). Next, the MZs in the second row are tuned sequentially so that all the power of the second code is routed to the second output of the MZM. All the rows in the MZM are tuned in a similar fashion. As an example, for the four-port MZM shown in Fig. 2, the MZIs are tuned in the order (row, column) $= (1,1) \rightarrow (1,2) \rightarrow (1,3) \rightarrow (1,4) \rightarrow (2,3) \rightarrow (2,4) \rightarrow (2,5) \rightarrow (3,5) \rightarrow (3,6)$. A single pass of self-configuration over all the rows configures the MZM, regardless of initially unknown phase shifts in the mesh due to temperature variations or manufacturing tolerances.

B. Analytical Framework

The N data signals can be expressed in vector (column) form as $\boldsymbol{u}(t) = [u_1(t), \ldots, u_N(t)]^\top$, where $^\top$ denotes vector transpose. Each data signal $u_m(t)$ is comprised of a high-speed on-off-keyed (OOK) information stream $d_m(t) \in \{0, 1\}$ that has been superimposed with a low-speed code $c_m(t) \in \{0, \pm 1\}$. We assume that each OOK information stream is randomly

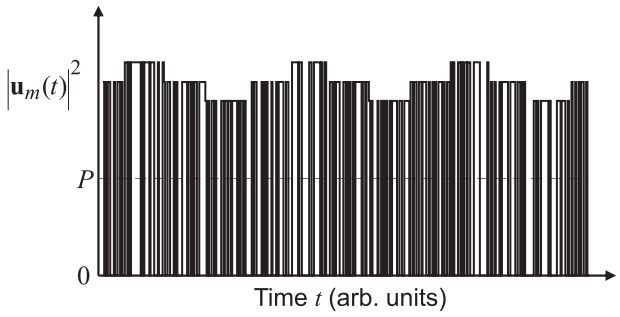

Fig. 3. The instantaneous power of an example data signal $\|\boldsymbol{u}_m(t)\|^2, 1 \leq m \leq N$, following the mathematical structure of (1). Each data signal $\boldsymbol{u}_m(t)$ is comprised of a high-speed OOK information stream $d_m(t)$ that is superimposed with a low-speed code $c_m(t)$ with a relative electric field modulation factor α. The low-speed codes modulate the power-envelope of the data signals. All data signals have an average power of P.

generated so $d_m(t)$ is 0 or 1 with equal probability. Although we discuss the exact form of the codes later, we assume that $c_m(t)$ is 0 half the time, $+1$ quarter of the time, and -1 quarter of the time. The form of each data signal can be expressed as

$$u_m(t) = e^{i\theta_m}\sqrt{\frac{4P}{2+\alpha^2}}\,(1+\alpha c_m(t))\,d_m(t), \quad \forall m = 1,\ldots,N, \tag{1}$$

where θ_m is the carrier phase of the data signal, which is assumed to vary much more slowly than the code and information sequences,[3] and the modulation factor $0 \leq \alpha \leq 1$ is the amplitude of the code electric field relative $c_m(t)$ to that of the information $d_m(t)$. We require $1+\alpha c_m(t) \geq 0$ to prevent the signal from becoming negative. Equation (1) contains a normalization factor such that the average power of the signal is P, independent of the chosen value of α. Note here we have chosen codes $c_m(t)$ that are $+1$, -1 or 0 at any given time, so we are not considering sinusoidal modulation as the codes. Rather, the codes are rectangular pulses and their design is discussed in Section III-C. The structure of the data signals is shown graphically in Fig. 3.

Consider the optical signal arriving at the input of the in-mesh PD corresponding to MZI (r,c), for some arbitrary row r and column c. Since the MZI couplers in the mesh are linear devices, the impinging electric field on this PD is a linear combination of all input data signals, given by $v_1^{(r,c)}(\Phi)u_1(t) + \cdots + v_N^{(r,c)}(\Phi)u_N(t) = \boldsymbol{v}^{(r,c)\top}(\Phi)\boldsymbol{u}(t)$, where $\boldsymbol{v}^{(r,c)}(\Phi)$ is an N-dimensional complex-valued vector that is a function of the waveguide channel and all the mesh phase shifter angles (in any preceding mesh elements) $\Phi = \{\theta_{av}^{(r,c)}, \Delta\theta^{(r,c)} | \text{ all valid } r,c\}$ (see Fig. 4 and Appendix VI). $\boldsymbol{v}^{(r,c)}(\Phi)$ is also a function of various losses in the mesh, either due to imperfect fabrication or due to losses from passing through in-mesh PDs. For notational convenience, we only write $\boldsymbol{v}^{(r,c)}$, dropping the explicit dependence on Φ hereafter. The self-configuration and related methods discussed in this paper separate the data signals by performing localized computations without assuming knowledge of $\boldsymbol{v}^{(r,c)}$, and so there is actually no need to explicitly estimate $\boldsymbol{v}^{(r,c)}$.

[3]If different lasers are used for each data signal, then the carrier phases of the data signals will be uncorrelated. Here, we assume that the light from a single laser is shared among all data signals, and so the carrier phases can be correlated. The methods presented in this article should work regardless of correlations in the carrier phases.

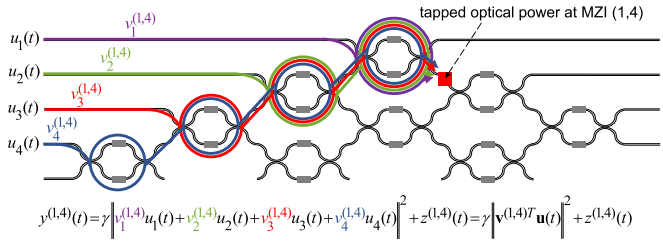

Fig. 4. Visualization of the $\boldsymbol{v}^{(r,c)}$ vectors for an MZM with $N=4$ input/output ports. For sake of simplicity, this figures assumes that there is no fiber channel prior to the MZM that causes mode mixing and the transmitted data signals $u_m(t)$ appear directly at the MZM input ports. As an example, (r,c) is taken to be (1,4) here. The total electric field at MZI (1,4) before being tapped at the in-mesh PD (1,4) is a linear combination of $u_1(t), u_2(t), u_3(t)$, and $u_4(t)$ with complex-valued weights $v_1^{(1,4)}, v_2^{(1,4)}, v_3^{(1,4)}$, and $v_4^{(1,4)}$, respectively. The complex scalar $v_m^{(1,4)}$ captures the losses and phase shifts seen by the mth data signal as it propagates through the MZM waveguides to in-mesh PD (1,4). Therefore, a noisy measurement of the tapped optical power can be expressed in vector dot-product notation as $y^{(1,4)}(t) = \gamma\|\boldsymbol{v}^{(1,4)\top}\boldsymbol{u}(t)\|^2 + z^{(1,4)}(t)$ where $(\cdot)^\top$ denotes vector transpose.

Assume that the in-mesh PDs tap a fraction γ of the optical power that impinge upon them. Then, the power detected at in-mesh PD (r,c) is

$$y^{(r,c)}(t) = \gamma\left\|\boldsymbol{v}^{(r,c)\top}\boldsymbol{u}(t)\right\|^2 + z^{(r,c)}(t), \tag{2}$$

where $\|\cdot\|^2$ is the squared modulus of a complex scalar, and $z^{(r,c)}(t)$ is additive thermal noise. In (2), $\boldsymbol{v}^{(r,c)}$ can be thought of as a weight vector for the contributions from each data signal. Since the information streams $d_k(t)$ are modelled as independent and identically distributed random OOK signals, it follows that the strength of the kth code $c_k(t)$, for $1 \leq k \leq N$, present in $y^{(r,c)}(t)$ is proportional to $\|v_k^{(r,c)}\|^2$.

Expanding (2), we get

$$\begin{aligned} y^{(r,c)}(t) &= \gamma\left\|\sum_{i=1}^{N} v_i^{(r,c)}u_i(t)\right\|^2 + z^{(r,c)}(t) \\ &= \frac{4\gamma P}{2+\alpha^2}\sum_{i,j=1}^{N}\Big\{v_i^{(r,c)}v_j^{(r,c)*}e^{j(\theta_i-\theta_j)}d_i(t)d_j(t) \\ &\quad\times\left[1+\alpha\left(c_i(t)+c_j(t)\right)+\alpha^2 c_i(t)c_j(t)\right]\Big\} + z^{(r,c)}(t). \end{aligned} \tag{3}$$

The squaring operation in direct detection mixes the various signal components, and so we observe in (3) that $y^{(r,c)}(t)$ is a sum of data-data cross products, code-data-data cross products, code-code-data-data cross products, and thermal noise. Adaptive methods must filter $y^{(r,c)}(t)$ to isolate signal components that contain information about the code strengths. The strength of the kth code, $1 \leq k \leq N$, can be extracted by filtering (3) using

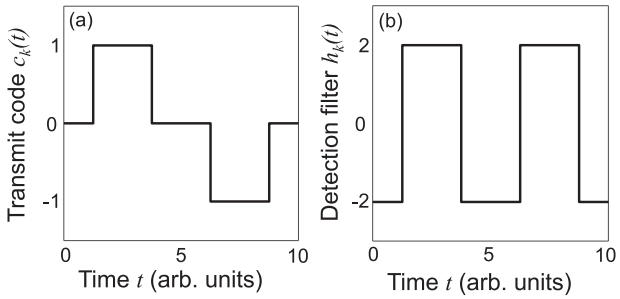

Fig. 5. An example (a) transmit code $c_k(t)$ and (b) corresponding detection filter $h_k(t)$ used for in-mesh detection of the code strengths, for some $1 \leq k \leq N$. The transmit codes and detection filters satisfy properties (5)–(8).

a detection filter $h_k(t)$ as

$$\mathcal{P}_k^{(r,c)} = h_k(t) * y^{(r,c)}(t) \propto \left\| v_k^{(r,c)} \right\|^2, \tag{4}$$

where the symbol $*$ denotes a filtering operation.[4] The following section describes the joint design of codes and detection filters.

C. In-Mesh Dectection of the Codes

We want to design a set of transmit codes $\{c_i(t) | 1 \leq i \leq N\}$ and corresponding detection filters $\{h_i^{\text{quad}}(t) | 1 \leq i \leq N\}$ that satisfy the properties

$$\mathbb{E}_t[h_i^{\text{quad}}(t)] = 0 \qquad \forall i \in 1, \ldots, N, \tag{5}$$

$$\mathbb{E}_t[h_i^{\text{quad}}(t) d_j(t)] = 0 \qquad \forall i, j \in 1, \ldots, N, \tag{6}$$

$$\mathbb{E}_t[h_i^{\text{quad}}(t) c_j(t)] = 0 \qquad \forall i, j \in 1, \ldots, N, \tag{7}$$

$$\mathbb{E}_t[h_i^{\text{quad}}(t) c_j(t) c_k(t)] = \delta_{ijk} \qquad \forall i, j, k \in 1, \ldots, N, \tag{8}$$

where $\mathbb{E}_t[\cdot]$ represents a time-averaging operator, and the Kronecker delta δ_{ijk} equals one when $i = j = k$ and zero otherwise. The superscript $(\cdot)^{\text{quad}}$ implies that filters $h_i^{\text{quad}}(t)$ are matched to the quadratic (i.e., squared) components of $c_i(t)$, because property (8) yields unity only when $h_i^{\text{quad}}(t)$ is correlated with $c_j(t)c_k(t)|_{i=j=k} = c_i^2(t)$.

One choice for the transmit codes that satisfy properties (5)–(8), while maximizing code power, are zero-mean bipolar square-wave codes with 50% duty-cycle return-to-zero pulses, chosen such that no frequency is the double of another and the sum of any two frequencies is not the double of another. The detection filters are then designed according to

$$h_i^{\text{quad}}(t) = 4\left(c_i(t)^2 - \mathbb{E}_t[c_i(t)^2]\right) \quad \forall i \in 1, \ldots, N. \tag{9}$$

An example transmit code and corresponding detection filter is shown in Fig. 5.

Suppose the self-configuration method calls for minimizing the power of some code $c_k(t)$ at the in-mesh PD (r, c). Equivalently, the phase shifter mesh angles Φ must be chosen such that

[4]For an arbitrary signal $g(t)$ and arbitrary filter $h(t)$, we define $h(t) * g(t) = \frac{1}{T}\int_0^T h(\tau) g(\tau) d\tau$ where $T = 1/\Delta f$ is the duration of filter h.

Method 1: Self-Configuration Method for Triangular MZMs with In-Mesh PDs.

```
for row r = 1 ... N − 1 do
  for column c ∈ valid columns given r, in ascending order do
    for tune step i = 1 ... N_tune do
      for angle θ ∈ (θ_av^(r,c), Δθ^(r,c)) do
        Minimize P_r^(r,c), the strength of code c_r at the bottom-right output of MZI (r, c), by adjusting θ. This maximizes the power at the top-right output of MZI (r, c).
      end for
    end for
  end for
end for
```

$\mathcal{P}_k^{(r,c)}$ is minimized. The mean value of the filter output is

$$\mathbb{E}_t\left[\mathcal{P}_k^{(r,c)}\right] = \frac{2\gamma\alpha^2 P}{2+\alpha^2} \cdot \underbrace{\left\| v_k^{(r,c)} \right\|^2}_{\text{desired component to minimize}}. \tag{10}$$

Notice that the measured code strength $\mathcal{P}_k^{(r,c)}$ is directly proportional to the strength of code $c_k(t)$ present in the total power detected at (r, c). This measurement is corrupted by additive interference and noise, and we have derived the variance of those terms in Appendix VI. The phase shifter mesh angles $\theta_{av}^{(r,c)}$ and $\Delta\theta^{(r,c)}$ can be set to minimize $\mathcal{P}_k^{(r,c)}$. In this paper, we assume that minimization is implemented in a sequence of N_{tune} steps. The minimization starts from a 2π search space for both angles, and a fixed fraction of the search space is eliminated at each tuning step by adjusting the lower and upper angle bounds. As a result, $\theta_{av}^{(r,c)}$ and $\Delta\theta^{(r,c)}$ converge exponentially to their optimal values after N_{tune} steps.[5]

The self-configuation method optimizes the phase shifts of all MZIs in a single pass using this procedure. The high-level intuitive explanation for the self-configuration method is provided in Section III-A, and the pseudocode for the method is provided in Method 1.

D. Performance of the Self-Configuration Method

In this section, we simulate the performance of a DD-MDM system using self-configuration of a triangular MZM with $N = 4$ input and output ports. Each of the N input data streams is modulated by OOK symbols at a sampling rate of $R_s = 25$ GBaud. The data streams are superimposed with low-power transmit code sequences using $\alpha = 0.1$, where α is the amplitude of the code electric field relative to that of the information, as defined in (1). The transmit codes are designed following the discussion in Section III-C. These signals propagate through a fully random unitary $N \times N$ MIMO channel prior to arriving at

[5]If the angle search space is reduced by a fixed fraction $f < 1$ during each iteration, then the optimal angle values lie in a search space of $2\pi \cdot f^{N_{\text{tune}}}$ after N_{tune} steps. In our simulations, f and N_{tune} are chosen such that the final search space is very small.

the inputs ports of the MZM. The MZM implements an $N \times N$ optical equalizer to undo modal crosstalk after propagation through the random channel. We assume that the code rate R_c is much smaller than R_s because channels in short-reach systems typically do not evolve faster than time scales of a few milliseconds [12], [13], and the MZM only needs to be updated in the sub-millisecond timescale.

Each in-mesh PD is assumed to be a p-i-n photodetector that taps a fraction γ of the incident power. The tapped powers at all in-mesh PDs are passed through narrowband code-detection filters, given by (9), in order to detect the strengths of the desired codes. The detected code strength is used to configure the MZM phase shifters, in accordance with the self-configuration process described in Method 1. The minimization at each MZI is performed with $N_{\text{tune}} = 10$ steps. The achieved bit-error ratio (BER) is calculated using an efficient semi-analytical enumeration method described in Appendix C.

The receiver sensitivity is found across a range of both received signal powers and MZM adaptation time. The total MZM adaptation time is equal to the total number of MZI phase shifters multiplied by the number of times each phase shifter is optimized (equal to one in self-configuration) and the time needed to adapt each MZI phase shifter (equal to N_{tune} multiplied by the code strength measurement time). We assume that code strengths are measured by a filtering operation at the system baud rate, which takes a finite amount of time. If the digital implementation of the code detection filters $h_k^{\text{quad}}(t)$ use N_{filt} taps, and the symbol rate is R_s, then each measurement of the code strength takes N_{filt}/R_s seconds. The MZM configuration time is varied by changing N_{filt}, and consequently the passband frequency width Δf, of the code detection filters. The received power of each data stream is varied from -25 dBm to -15 dBm, and at each test point, N_{filt} is chosen between 2^{20} and 2^{32} taps such that the achieved average BER is less than a target forward-error-correction (FEC) threshold of 1.8×10^{-4}. This means that each measurement of the code strength can take between 42 μs and 172 ms, depending on the value of N_{filt} used. For detecting low speed signals below 10 MHz we assumed TIAs with noise levels of 3 pA/$\sqrt{\text{Hz}}$, while for higher-frequency signals we assumed higher noise levels ranging up to 30 pA/$\sqrt{\text{Hz}}$ at 10 GHz, following the cube-root of the required TIA bandwidth. Once the code strengths are measured, we assume that the calculations needed to find the optimal MZI phase shifter angles are instantaneous (even though, in practice, they would also take a finite amount of time to compute in a hardware DSP or FPGA). Therefore, the total MZM adaptation times presented in this paper are a lower bound for a realistic implementation, but they still provide insightful information because they are independent of any specific computer, software (MATLAB or compiled code) or used routines. A table of the key parameters we assumed in our simulations is presented in Table I.

Fig. 6 shows the impact of the in-mesh power tap-off ratio γ at each MZI on the achieved per-mode signal-to-noise ratio (SNR) after the MZM has been configured. When γ is small, ranging from 0% to 2%, the tapped optical power at each MZI is also small, and thermal noise added by the in-mesh PDs causes the achieved per-mode SNR to be too low. When γ is large ($> 10\%$), the MZIs closer to the MZM input ports absorb too much of the transmitted optical power, which hinders the code detection process for the MZIs in the later rows. This consequently leads to a decrease in the achieved per-mode SNRs. The optimal value of γ is found to be close to 3%, which is a judicious tradeoff between the two extreme regimes.

Fig. 7 shows the achieved receiver sensitivity for the self-configuration method, and also for the methods discussed later in Section IV. It is apparent that the self-configuration method in its original form is too slow to reliably configure the MZM in a

TABLE I
SIMULATION PARAMETERS FOR EVALUATING MZM ADAPTATION METHODS

Parameter	Symbol	Value
Number of MZM input/output ports	N	4
Symbol rate	R_s	25 Gbaud
Code rate	R_c	$\ll R_s$
Relative electric field amplitude of codes	α	0.1
Optical power per mode	P	-15 to -25 dBm
Information modulation format		OOK
Number of taps of detection filters	N_{filt}	varied from 2^{20} to 2^{32}
Time to detect codes at each PD	T_{filt}	N_{filt}/R_s
Tuning steps in angle search	N_{tune}	10
In-mesh p-i-n PD tap-off ratio	γ	3%
Output detection MZM iterations		35
Curve fitting test angles		20
Responsitivity	R	1 A/W (p-i-n PD)
		0.74 A/W (APD)
APD low-gain bandwidth		24 GHz
APD gain-bandwidth product		290 GHz
APD multiplication ratio	G	10
APD impact ionization coefficients ratio	k_a	0.18
Dark current	I_d	40 nA
TIA noise figure	F_n	1.5
TIA feedback resistor	R_F	30 Ω (APD)
		30-3000 Ω (p-i-n PD)
TIA noise level		30 pA/$\sqrt{\text{Hz}}$ (APD)
		3-30 pA/$\sqrt{\text{Hz}}$ (p-i-n PD)
Absolute temperature	T	300 K

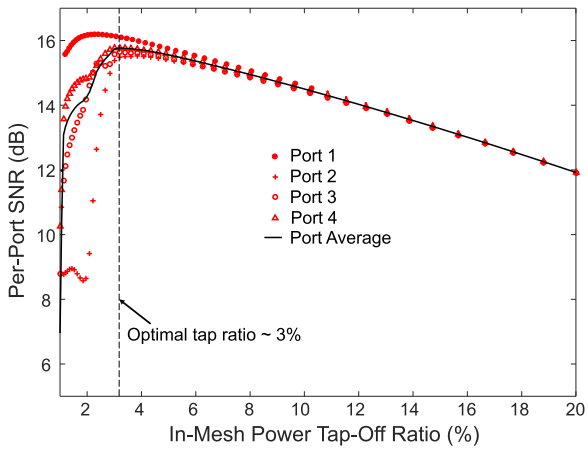

Fig. 6. Optimizing the power tap ratio γ at each in-mesh PD of a triangular MZM with $N = 4$ input and output ports. The per-port SNRs after adapting the MZM are plotted as a function of γ. At each test point, the MZM adaptation time is held constant by fixing the value of $N_{\text{filt}} = 2^{20}$ taps. The optimal γ that maximizes the port-averaged SNR is found to be close to 3%.

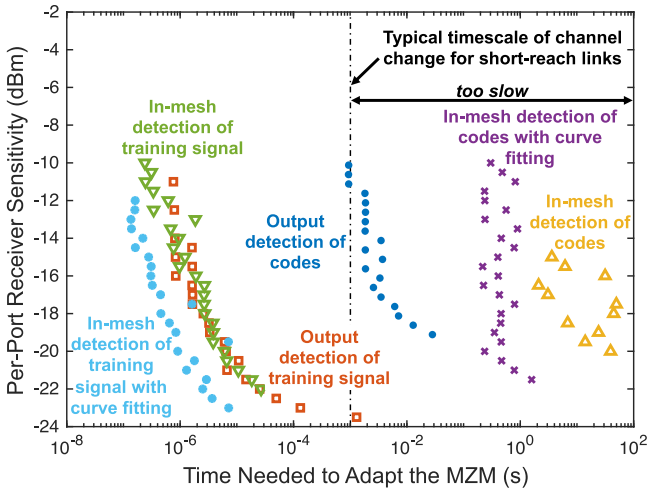

Fig. 7. Per-port receiver sensitivity achieved after adapting a triangular MZM with $N = 4$ input and output ports using various methods. The time needed to adapt the MZM is varied by changing the number of taps N_{filt} in the code detection filters (for methods using codes) or by changing the duration of training signals T_{train} (for methods using full-power training sequences). In-mesh detection of codes (shown in the figure using yellow triangles) is equivalent to the self-configuration method discussed in Section III. The remaining methods are discussed further in Section IV. Channels of short-reach links typically evolve on the order of milliseconds, so methods that take much longer to adapt the MZM may not reliably work.

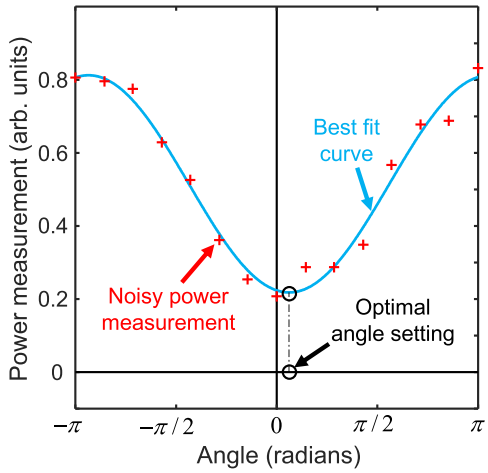

Fig. 8. Illustration of the curve fitting method. A function of the form $\|p_1 \sin(p_2\theta + p_3)\|^2 + p_4$, shown in the plot as a solid blue curve, was fit to noisy power measurements, shown by red plus signs, for several values of θ in the range $[0, 2\pi]$. This function can be analytically minimized to find the optimal angle.

practical data transmission setting. This result can be explained as follows. Since the self-configuration method passes over each MZI exactly once, accurate estimates of the relevant code strengths are needed to find the optimal phase shifts. However, as suggested by the prefactor and cross-terms in (3), the SNR of the tapped optical power is small. Therefore, the filtering operations given by (10) need to have long durations to reduce the impact of noise and produce reliable estimates of the relevant code strengths. As seen by the yellow triangle markers in Fig. 7, our simulations indicate that a single configuration pass through all the MZIs of the MZM typically takes tens of seconds to achieve the necessary BER threshold. Although our simulations assume that the MZM phase shifters are instantaneous, this result is consistent with experimental observations in [11], where a full reconfiguration of a four-port triangular MZM using thermo-optic phase shifters took approximately 15 seconds. In the next section, we present small modifications to the self-configuration method that enable the MZM to converge and track considerably faster channel dynamics.

IV. Improving the Self-Configuration Method

We propose increasingly faster modifications to the self-configuration method in each of the following sections. An overlay plot comparing the performance of all MZM configuration methods presented in this paper is shown in Fig. 7.

A. In-Mesh Detection of Codes With Curve Fitting

Curve fitting is a simple optimization trick that can potentially allow finding the optimal MZM phase shifter angles with fewer measurements. In the original self-configuration method of Section III, each MZI was optimized by a brute-force angle search procedure. Convergence to the optimal phase shifter angles can require a high number of search steps when the SNR is low. Here, we leverage the observation that the detected optical power at MZI (r, c), $\mathcal{P}_k^{(r,c)}$, is a simple function of the MZI phase shifts $\theta_{av}^{(r,c)}$ and $\Delta\theta^{(r,c)}$. When one of these two angles is fixed and the other is allowed to vary as θ, $\mathcal{P}_k^{(r,c)}$ takes the form

$$\mathcal{P}_k^{(r,c)} = \|p_1 \sin(p_2\theta + p_3)\|^2 + p_4, \tag{11}$$

for some constants p_1, p_2, p_3, p_4.

A nonlinear curve fitting procedure, e.g. based on the Levenberg-Marquardt algorithm [26]–[28], can fit p_i to power measurements at a sequence of discrete angles spaced from 0 to 2π radians. The resulting function can be analytically minimized to yield the optimal phase shift θ. Alternating minimization over $\theta_{av}^{(r,c)}$ and $\Delta\theta^{(r,c)}$ yields the optimal phase shifts for each MZI in the MZM. An illustration of nonlinear curve fitting is shown in Fig. 8. In evaluating the performance of the curve fitting method, we used the same simulation setup described in Section III-D and assumed the key parameters from Table I. We used 10 different power measurements at discrete phase values per phase control and used MATLAB's Curve Fitting toolbox to perform the nonlinear curve fit. Curve fitting reduces the impact of measurement noise because all power measurements contribute to the best analytic fit, causing the per-sample measurement noise to be averaged across all samples. As seen from the purple cross markers in Fig. 7, this method typically provides an overall speed-up factor of roughly ten, as compared to the brute-force minimization angle search of Section III. However, in-mesh detection of codes with curve fitting may still be too slow for practical data transmission, as this method adapts the MZM in roughly 1 s.

Method 2: Adaptation Method for Triangular MZMs with PDs at the Output Ports Only.

for output port (or equivalently, row) $n = 1 \dots N-1$ **do**
  for iteration $i = 1 \dots N_{iter}$ **do**
    Let $\mathcal{A}$ be the set of common and differential-mode phase controls for each MZI in row n
    for each phase control in $\mathcal{A}$ **do**
      Measure strength of code c_n on each output port as a vector: $\mathcal{P}_n = [\mathcal{P}_n^1, \dots, \mathcal{P}_n^N]^\top$.
      Compute error $\|(\mathcal{P}_n/\|\mathcal{P}_n\|) - \mathbf{e}_n\|^2$ where $\mathbf{e}_n$ is one in the nth element and zero elsewhere.
      Measure the error at S different phase adjustments and set the phase control to the value that yields the lowest error. The search space can reduce exponentially with each iteration i.
    end for
  end for
end for

Curve fitting to a sinusoidal response to a 2π range of MZM phase angles is also an idealized mechanism to test convergence speed. When adapting a live communication system to track changing fiber mode dynamics, testing a full range of MZM phases will interfere with the live data signals. Curve fitting to a small range of phase adjustments would be necessary to create a viable communication system, if the tracking of codes modulated on top of data signals enabled fast enough tracking.

B. Output Detection of Codes

The original self-configuration method optimizes each MZI coupler, in sequence, using a *single* pass through the MZM. It takes a long time to optimize each MZI coupler because the codes are measured at a low SNR at each in-mesh PD. In this method, we propose an alternative MZM adaptation method that forgoes in-mesh PDs and uses measurements from APDs at the output ports instead. Since the entire optical power of the data signals is being detected at the output ports, the code strengths are detected at higher SNRs. However, as we explain below, these code strength measurements are corrupted by modal interference. Output detection of codes is an iterative method. We loop through the MZM *multiple* times, and at each iteration, set the MZI phase shifters to approximately optimal values based on power measurements at the output APDs. The impact of the modal interference reduces with each iteration, and the MZM converges after a sufficient number of iterations. Even though more iterations through the MZM are required than the self-configuration method, we show that higher measurement SNRs provided by output detection lead to a significant increase in adaptation speed compared to the self-configuration method.

Data signals optimized for this method have the form

$$u_m(t) = e^{i\theta_m}\sqrt{\frac{2P}{1+\alpha^2}}\left(1+\alpha c_m(t)\right)d_m(t), \quad \forall m = 1,\dots,N. \tag{12}$$

The normalization factor assumes that the OOK information streams are generated randomly (i.e., $d_m(t)$ is 0 or 1 with equal probability), and $c_m(t)$ is $+1$ or -1 with equal probability. Note that (12) differs from (1) because the optimized codes for output detection do not need a zero level. Similar to (2), the total optical power detected by the APD at output port n is

$$y^{(n)}(t) = \left\|\boldsymbol{v}^{(n)\top}\boldsymbol{u}(t)\right\|^2 + z^{(n)}(t), \tag{13}$$

where $\boldsymbol{v}^{(n)}$ is a N-dimensional vector that linearly weights the contributions of the data signals at the nth output port, and $z^{(n)}(t)$ is the shot noise added by the APD. Note that $\gamma = 1$ because the full optical power is detected, instead of a small tapped fraction.

In order to adapt the MZM so that it routes each $u_m(t)$, $1 \le m \le N$, to its corresponding output port, we design transmit code sequences $\{c_i(t)|1 \le i \le N\}$ and detection filters $\{h_i^{\text{linear}}(t)|1 \le i \le N\}$ that satisfy constraints given by

$$\mathbb{E}_t[h_i^{\text{linear}}(t)] = 0 \qquad \forall i \in 1,\dots,N, \tag{14}$$

$$\mathbb{E}_t[h_i^{\text{linear}}(t)c_j(t)] = \delta_{ij} \qquad \forall i,j \in 1,\dots,N, \tag{15}$$

$$\mathbb{E}_t[h_i^{\text{linear}}(t)d_j(t)] = 0 \qquad \forall i,j \in 1,\dots,N, \tag{16}$$

$$\mathbb{E}_t[h_i^{\text{linear}}(t)c_j(t)c_k(t)] = 0 \qquad \forall i,j,k \in 1,\dots,N. \tag{17}$$

The detection filters for this method are given by

$$h_k^{\text{linear}}(t) = c_k(t), \tag{18}$$

where the superscript $(\cdot)^{\text{linear}}$ implies that the filters $h_i^{\text{linear}}(t)$ are matched to (or correlated with) components of $c_i(t)$. Property (14) implies that the codes are zero-mean and property (15) implies that they are mutually orthogonal. Property (16) implies that the code and information sequences are uncorrelated, and finally, property (17) implies that the codes are orthogonal to any pairwise multiplication. A set of codes that satisfy these properties are the even rows of Hadamard matrices for dimensions that are powers of 2, formed via Sylvester's condition [29]. MATLAB's `hadamard()` function automatically follows Sylvester's construction for dimensions that are powers of 2. An example set of linear codes which satisfy all conditions (14)–(17) is shown in Fig. 9 for $N = 4$.

After the detection filters are applied, the mean power of code k at output port n becomes

$$\mathcal{P}_k^n = \frac{2\alpha P}{1+\alpha^2}\left[\underbrace{\left\|v_k^{(n)}\right\|^2}_{\text{desired component}} + \cdots \underbrace{\sum_{i=1,i\neq k}^{N}\text{Re}\left(v_i^{(n)}v_k^{(n)*}e^{j(\theta_i-\theta_k)}\right)}_{\text{modal interference}}\right]. \tag{19}$$

Note that (19) shows that the output port power measurements are corrupted by the modal interference terms. It is not possible to choose the MZM phase shifts to minimize the desired components in a single pass through the mesh. Each MZI coupler

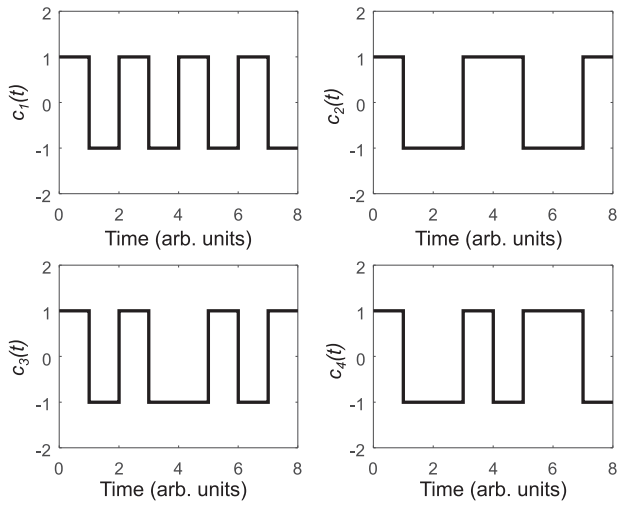

Fig. 9. A set of $N = 4$ linear codes formed from the even rows of the Hadamard matrix of size 8. These codes satisfy all the properties (14)–(17).

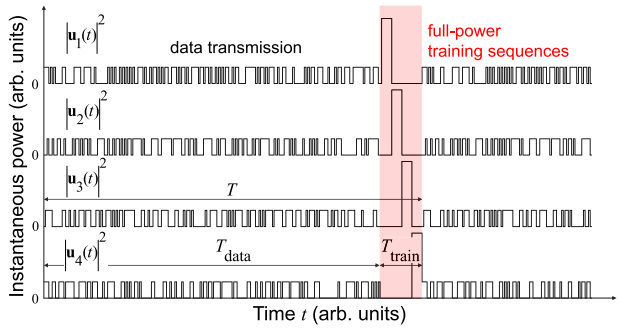

Fig. 10. Instantaneous power of data signals $u_m(t), m = 1, \ldots, N$. Data transmission is periodically stopped every $T = T_{\text{train}} + T_{\text{data}}$ seconds and each of the data signals is instead dedicated to transmitting a full-power training signal to aid in adapting the MZM. Example power traces are shown for the case of an MZM with $N = 4$ input and output ports. The training signals are continuous-wave pulses of duration T_{train}/N seconds that are time-multiplexed into a window of duration T_{train}. Since they do not overlap in time, the training pulses have power $P \cdot N$ where P is the average power of each signal during the data transmission phases.

must be optimized multiple times, and in each iteration, the impact of the modal interference terms reduces. With enough iterations, the MZM can be configured to the channel inverse. The pseudocode for this method based on output port detection of codes is summarized in Method 2. Simulations of this method using the key parameters from Table I show that output detection of codes can converge on the order of 10^{-2} seconds, as observed by the dark blue dot markers in Fig. 7.

C. Output Detection of Training Signals

All of the MZM tuning methods proposed thus far continuously transmit low-power code sequences that are superimposed onto high-speed information streams, in order to continuously adapt the MZM to the link's channel dynamics. In this method, we propose a *periodic* equalization scheme. As shown in Fig. 10, each data signal is comprised of packets of length $T = T_{\text{train}} + T_{\text{data}}$ seconds. The first T_{train} seconds of each packet is dedicated to adapting the MZM using full-power training signals. We use a simple time-division multiplexing (TDM) scheme here, where continuous-wave pulses of power $P \cdot N$ and duration T_{train}/N seconds are sent one after another on the respective data signals during a time window of duration T_{train} seconds.

Power measurements from the output ports are used to set the MZM phase controls, and multiple passes through the MZM are required before the MZI phase shifts converge to their optimal values. This method uses similar steps as Method 2. As seen in Fig. 7, we observe a significant speedup in the time needed for adjusting the MZM. This is because the TDM scheme assures there is no interference (similar in kind to modal interference from Section IV-B) from parallel training signals. Since high-speed information transmission is stopped when the MZM is being adapted, there is also no data interference. The peak power of each training pulse can be N times higher than the average power of each data signal without increasing the average power, since only one training pulse is active at any given time. This contributes to a further increase in the measurement SNRs at the output PDs.

Periodic equalization leads to a small, albeit insignificant, decrease in information throughput because the MZM can be adapted very quickly between bursts of information transmission. Fig. 7 suggests that this method can adapt the MZM in timescales close to 10^{-5} seconds, and even assuming the channel changes every 10^{-3} seconds, the training overhead is as small as $T_{\text{train}}/T_{\text{data}} \approx 1\%$. Most of the transmission packet can be spent delivering payload information streams to the receiver, during the T_{data} phase.

D. In-Mesh Detection of Training Signals With Curve Fitting

The fastest adaptation method that we present in this paper periodically detects a tapped portion of full-power training signals at each in-mesh PD of the MZM. Similar to Section IV-C, the training signals can be continuous-wave pulses that are time-multiplexed into a single window of duration T_{train}. The steps of the adaptation method are similar in essence to those described in Method 1, except that we detect the power of continous-wave pulses at each in-mesh PD instead of codes, and curve-fitting can be used to find the optimal MZI phase shifts instead of a brute-force search. This method performs exceedingly well for several reasons:

1) High measurement SNRs: although a small fraction γ of the incident optical power is detected at each in-mesh PD, the training signal pulses have a power $P \cdot N$, which is high enough to accurately set the MZI phase shifts even in the presence of thermal noise.
2) No data interference: since high-speed information is not transmitted at the same time as the low-speed training signals, there is no data interference. Interference from other training signals is also avoided because the training signals are mutually orthogonal in time.
3) Sequential minimization of each MZI: Similar to the original self-configuration method, this method requires a single pass through the MZM to configure all the MZI

phase shifts. Multiple passes through the MZM are not required, as in the case of output detection (Sections IV-B and IV-C).

4) Curve-fitting: As discussed in Section IV-A, setting the MZI phase shifts by minimizing an analytic function reduces the number of discrete phase values to check. Furthermore, the inherent averaging that occurs during the curve-fitting process reduces the impact of per-sample measurement noise.

Our simulations show that MZMs can be configured in microsecond timescales using this method. These optimized methods (green triangle and light-blue dot markers) are roughly 10^8-fold faster than the original self-configuration method of Section III (yellow triangle markers in Fig. 7).

V. Discussion

The requirements of short-reach DD-MDM systems, which can potentially be deployed in data centers to increase per-fiber information capacity, are stringent because they must operate under tight power and cost budget constraints. Experiments have shown that the channels in short-reach links change on time scales of the order of milliseconds, and that the speed of channel dynamics increases with mode count if the modes are coupled [14]. Although we have presented several equalization methods for adapting optical MIMO hardware using triangular MZMs, not all of them can track dynamic MIMO channels with sufficient speed. As seen in Fig. 7, methods using low-power and low-speed code signals, such as in-mesh detection of codes (with or without curve fitting) and output detection of codes, can require 10^{-2} or more seconds even to adapt a relatively small triangular MZM with $N = 4$ input and output ports. Initial simulations show that the MZM adaptation time scales slightly above quadratically with N. These methods are too slow to be deployed during high-speed transmission because the receiver can experience outages if the channel changes faster than the optical equalizer. By contrast, periodic equalization methods using full-power training signals are better suited for tracking DD-MDM links.

In this work, we have made the simplifying assumption that the MZI phase shifts can be changed instantaneously. Of course, this does not hold true in practice because the phase shifts are typically controlled by integrated heaters in silicon photonics systems, and the time responses of thermo-optic actuators are in the order of 10 μs [30], [31]. The speed of MZM tracking may be limited by the speed of the phase shifting elements once the optimized methods relying on full-power training sequences are used. Recent work shows that it is possible to build MZ-based CMOS photonic switch fabrics with nanosecond-scale switching speed [32]. MEMS-actuated silicon photonics can also be a promising avenue for configuring MZMs in the microsecond timescale [33]. Furthermore, it is possible that burst errors can occur during MZM tracking when a MZI phase shifter has to be changed by a large value, similar to the problems typically associated with phase resets and endless tracking using analog control loops. This can be avoided by changing the MZI phase shifters only during predefined equalization timeslots when data transmission is turned off (i.e., by using full-power training sequences), or by using interleaving and correcting burst errors using FEC [34].

In-mesh PDs require complicated control circuitry to measure a small fraction of the power in the waveguide and to detect the code strengths using filtering operations. Since the number of MZIs per MZM scales quadratically with the number of input and output ports N, it may become prohibitively expensive to build low-cost hardware for large N. Our simulations show that it is possible to adapt MZMs sufficiently well using APDs at the output ports alone. The hardware complexity for output APDs scales linearly with N, which may be more manageable.

Future directions for this research include studying the performance of high-order modulation formats such as four-level pulse-amplitude-modulation, studying the impact of component bandwith limitations, and studying group delay compensation methods to equalize frequency-dependent effects in longer DD-MDM links.

VI. Conclusion

In this paper, we studied the problem of adjusting phase shifters inside a triangular MZM to perform optical MIMO equalization in short-reach DD-MDM systems. In contrast to coherent detection systems where equalization can be performed digitally using measurements of the received electric field, DD-MDM systems must equalize signals in the optical domain using received intensity measurements. We studied the performance of the self-configuration method by simulating a link with $D = 4$ data signals. Fig. 7 shows that impairments from data interference and noise only allow the MZM to be adjusted in tens of seconds and therefore limit its tracking ability. We proposed several variants of the self-configuration method by considering the placement of PDs in the MZM and by analyzing the tradeoffs between continuous tracking using low-power code sequences and periodic tracking using full-power training signals. Our analysis shows that the biggest gains in MZM tracking ability are gained by periodically adjusting the mesh using high-power training signals because equalization is performed at high SNRs. MZM architectures using either in-mesh PDs or output PDs may work equally well if their adaptation methods are designed accordingly. The fastest adaptation methods that we have studied in this paper allow the MZM to track channel dynamics on the order of 10^{-6} seconds.

Appendix A
Variance of the Intereference and Noise Terms in the Self-Configuration Method

We derive expressions for the interference and noise variances that impair MZM adaptation, which were used in the simulations of Section III. Although we focus only on the self-configuration method here, expressions for the alternate methods in Section IV can be derived using similar means.

Suppose we are interested in measuring the strength of code $c_k(t)$ at the in-mesh PD (r, c). The result of filtering the total detected power at in-mesh PD (r, c), $y^{(r,c)}(t)$, with a detection

filter $h_k^{\text{quad}}(t)$ is

$$\begin{aligned}\mathcal{P}_k^{(r,c)} &= h_k^{\text{quad}}(t) * y^{(r,c)}(t) \\ &= \frac{4\gamma P}{2+\alpha^2} \sum_{i=1}^{N}\sum_{j=1}^{N} \Big\{ v_i^{(r,c)} v_j^{(r,c)*} e^{j(\theta_i-\theta_j)} \\ &\quad \times \Big(h_k^{\text{quad}}(t) * d_i(t) d_j(t) && \text{term } A_{ijk} \\ &\quad + \alpha h_k^{\text{quad}}(t) * d_i(t) d_j(t) \left[c_i(t) + c_j(t)\right] && \text{term } B_{ijk} \\ &\quad + \alpha^2 h_k^{\text{quad}}(t) * d_i(t) d_j(t) c_i(t) c_j(t) \Big) \Big\} && \text{term } C_{ijk} \\ &\quad + h_k^{\text{quad}}(t) * z^{(r,c)}(t). && \text{term } D_{ijk} \end{aligned} \quad (20)$$

Terms A_{ijk} and B_{ijk} contribute to *data interference* resulting from filtering the cross products of high-speed information signals $d_i(t)$. Term D_{ijk} results from filtering the *thermal noise* $z^{(r,c)}(t)$ (with noise variance σ_{th}^2). Term C_{ijk} is the component we wish to detect because $h_k^{\text{quad}}(t)$ was designed to correlate against $c_k^2(t)$.

A full simulation of the self-configuration method would require us to generate high-speed information streams $d_i(t)$ to quantify the impact of terms A_{ijk}, B_{ijk}, C_{ijk} and D_{ijk}. However, we note that $d_i(t)$ are independent stochastic signals, and that these terms can be approximated as additive Gaussian random variables with means and variances given by analytical formulas. The formulas allow us to bypass full simulations that use high-speed information streams, thereby significantly speeding up simulations.

Suppose that the codes $c_i(t)$ and detection filters $h_i^{\text{quad}}(t)$ have been jointly designed to satisfy (5)-(8). The information streams $d_i(t)$ can be modelled as independent stochastic signals from the set $\{0, +1\}$ with equal probability. The codes can be modelled as independent discrete random variables which assume values $+1, -1, 0$ with probabilities $1/4, 1/4$ and $1/2$, respectively. The detection filters are matched to (or correlated with) $c_i(t)^2$ and assume a value of $+2$ when $c_i(t) = +1$ or -1, and a value of -2 when $c_i(t) = 0$. Also, the detection filters are narrowband filters with a passband frequency width Δf. In our simulations, we implement the codes and filters with arrays of length N_{filt}, and so $\Delta f = 1/N_{\text{filt}}$.

Using the central limit theorem, we can replace terms A_{ijk}, B_{ijk}, C_{ijk}, and D_{ijk} with random Gaussian variables. By a brute-force enumeration over all possible states of $d_i(t), d_j(t), c_i(t), c_j(t)$ and $h_i^{\text{quad}}(t)$, we can show that

$$\text{term } A_{ijk} \sim \mathcal{N}\left(0, (1+\delta_{ij})\Delta f\right), \quad (21)$$

$$\text{term } B_{ijk} \sim \mathcal{N}\left(0, \alpha^2(1+3\delta_{ij})\Delta f\right), \quad (22)$$

$$\text{term } C_{ijk} \sim \mathcal{N}\left(\frac{\alpha^2\delta_{ijk}}{2}, \frac{\alpha^4(1+3\delta_{ij}-\delta_{ijk})\Delta f}{4}\right), \quad (23)$$

$$\text{term } D_{ijk} \sim \mathcal{N}\left(0, \sigma_{\text{th}}^2 \Delta f\right), \quad (24)$$

where $\mathcal{N}(\mu, \sigma^2)$ is the standard normal (or Gaussian) distribution with mean μ and variance σ^2. Using (21)–(24), the mean value of (20) becomes

$$\mathbb{E}_t\left[\mathcal{P}_k^{(r,c)}\right] = \frac{2\gamma\alpha^2 P}{2+\alpha^2} \cdot \left\| v_k^{(r,c)} \right\|^2, \quad (25)$$

which is proportional to the strength of the kth code, in agreement with (10). We can observe that increasing the integration time $1/\Delta f = N_{\text{filt}}$ of the filter output decreases the variance of the interference and noise terms, similar to the effect observed in averaging filters. This poses a tradeoff: we desire to choose Δf small enough such noise is reduced, but large enough to enable fast adaptation of the mesh. Note also that the intereference and noise terms in (20) are actually time-varying because the $v_i^{(r,c)} v_j^{(r,c)*}$ in the sum depend on the MZM phase shifts, which themselves change with time as the self-configuration method progresses through each MZI coupler.

Appendix B
MZM Transmission Matrices

The triangular MZM with N input and output ports for perfoming $N \times N$ optical MIMO functions is shown in Fig. 2. The building block of the MZM, the MZI coupler, is shown in the inset and consists of an ideal 50% coupler, followed by two arms with phase shifting elements, and followed by another ideal 50% coupler. By applying voltages to the phase shifting arms, the common-mode phase shift θ_{av} and the differential phase shift $\Delta\theta$ can be controlled. The 2×2 transmission matrix of the MZI in row r and column c is given by

$$\begin{aligned}\mathbf{H}_{\text{MZI}}^{(r,c)}\left(\theta_{av}^{(r,c)}, \Delta\theta^{(r,c)}\right) &= \frac{1}{2}\begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \times \cdots \\ &\begin{bmatrix} e^{j\left(\theta_{av}^{(r,c)} + \Delta\theta^{(r,c)}/2 + \theta_{cal,1}^{(r,c)}\right)} & 0 \\ 0 & e^{j\left(\theta_{av}^{(r,c)} - \Delta\theta^{(r,c)}/2 + \theta_{cal,2}^{(r,c)}\right)} \end{bmatrix} \\ &\times \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \end{aligned} \quad (26)$$

where $\theta_{cal,1}^{(r,c)}, \theta_{cal,2}^{(r,c)}$ are the unknown phase shifts of the arms due to manufacturing tolerances [1]. In our simulations, we assume that $\theta_{cal,1}^{(r,c)}, \theta_{cal,2}^{(r,c)}$ are random variables distributed uniformly in the range $[0, 2\pi]$. By appropriate control of $\theta_{av}^{(r,c)}$ and $\Delta\theta^{(r,c)}$, any 2×2 unitary transmission matrix can be physically realized. By controlling the phase shifts in all MZI couplers, as discussed in [1] and Sections III and IV, any $N \times N$ unitary transmission matrix can be constructed.

Appendix C
Semi-Analytic Calculation of Bit-Error Ratio

We discuss how to efficiently calculate the average BER of the link for any particular configuration of the MZM phase shifters. Suppose that the fiber supports K mode groups and has transmission matrix $\mathbf{H}_{\text{fiber}}$. The transmission matrix of each MZM (one MZM for each mode group) is given by $\mathbf{H}_{\text{MZM}}^{(i)}, 1 \le i \le K$. The combined transmission matrix of the fiber channel and the

MZM equalizers is given by the $D \times D$ matrix

$$\mathbf{H} = \mathbf{H}_{\mathrm{MZM}}^{(\mathrm{tot})} \mathbf{H}_{\mathrm{fiber}}, \tag{27}$$

where $\mathbf{H}_{\mathrm{MZM}}^{(\mathrm{tot})}$ is a block diagonal matrix of the form

$$\mathbf{H}_{\mathrm{MZM}}^{(\mathrm{tot})} = \begin{bmatrix} \mathbf{H}_{\mathrm{MZM}}^{(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{H}_{\mathrm{MZM}}^{(K)} \end{bmatrix}. \tag{28}$$

Suppose also that the transmitted symbols are given by the $D \times 1$ vector $\mathbf{u}$. Looking at the structure of (1), each element of $\mathbf{u}$, given by u_i for $1 \leq i \leq D$, can assume four different amplitudes: one amplitude for when the data $d_i = 0$ (which occurs with probability $p(u_i) = 1/2$), and three amplitudes for when the data $d_i = 1$ (corresponding to the three levels of the code c_i, and each occurs with probability $p(u_i) = 1/6$). Therefore, $\mathbf{u}$ can assume 4^D discrete vectors, each with probability $p(\mathbf{u}) = \prod_{i=1}^{D} p(u_i)$.

We also define an *interference* matrix, given by

$$\tilde{\mathbf{H}} = \mathbf{H} - \mathrm{diag}\left(\mathbf{H}\right), \tag{29}$$

where $\mathrm{diag}(\mathbf{H})$ is a diagonal matrix constructed from the elements on the main diagonal of $\mathbf{H}$.

For each value of $\mathbf{u}$, we calculate:

1) $\mathbf{p}_0(\mathbf{u}) = [\mathbf{p}_{01}(\mathbf{u}), \ldots, \mathbf{p}_{0D}(\mathbf{u})]^\top$: a $D \times 1$ vector of received interference powers at the output APDs when each transmitted signal is turned off, one-by-one, while the remaining transmitted signals are set according to $\mathbf{u}$. Its elements are defined as

$$\mathbf{p}_{0i}(\mathbf{u}) = \left\| \tilde{\mathbf{H}}_{i:}^{\top} \mathbf{u} \right\|^2, \tag{30}$$

where $\tilde{\mathbf{H}}_{i:}$ is the ith row of $\tilde{\mathbf{H}}$.

2) $\mathbf{p}_1(\mathbf{u}) = [\mathbf{p}_{11}(\mathbf{u}), \ldots, \mathbf{p}_{1D}(\mathbf{u})]^\top$: a $D \times 1$ vector of received signal powers at each output APD when each transmitted signal is turned on, one-by-one, while the remaining transmitted signals are set according to $\mathbf{u}$. Its elements are given by

$$\mathbf{p}_{1i}(\mathbf{u}) = \left\| \mathbf{H}_{i:}^{\top} \mathbf{u}^{(i)} \right\|^2, \tag{31}$$

where $\mathbf{u}^{(i)}$ is identical to $\mathbf{u}$, except in the ith position where it is set to encode a logical value of 1.

3) $\boldsymbol{\sigma}_{n,0}^2(\mathbf{u})$ and $\sigma_{n,1}^2(\mathbf{u})$: vectors of noise variances. They are derived as

$$\sigma_{n,q}^2(\mathbf{u}) = F_{n,\mathrm{eq}} \cdot \left(\sigma_{\mathrm{th}}^2 + \sigma_{\mathrm{shot},q}^2(\mathbf{u})\right), \tag{32}$$

$$\sigma_{\mathrm{th}}^2 = 4k_B T F_n \Delta f_{\mathrm{APD}} / R_{F,\mathrm{APD}}, \tag{33}$$

$$\sigma_{\mathrm{shot},q}^2(\mathbf{u}) = 2qG^2 F_A \left(R_{\mathrm{APD}} + \mathbf{p}_q(\mathbf{u}) + I_d\right) \Delta f_{\mathrm{APD}}, \tag{34}$$

where $F_{n,\mathrm{eq}}$ is a penalty factor for equalization noise enhancement, σ_{th}^2 is the signal-independent thermal noise variance, $\sigma_{\mathrm{shot},q}^2$ is the signal-dependent shot noise for logical level $q \in \{0, 1\}$, F_n is the TIA noise figure, T is the absolute temperature, k_B is Boltzmann's constant, $R_{\mathrm{F,APD}}$ is the TIA feedback resistor, Δf_{APD} is the APD bandwidth, R_{APD} is the APD responsitivity, q is the elementary charge, and G is the APD gain or multiplication ratio. The excess noise factor of the APD F_A is given by

$$F_A = k_a G + (1 - k_a)\left(2 - \frac{1}{G}\right). \tag{35}$$

The relevant parameters of the output APDs are defined in Table I.

Once the above quantities are computed for all $\mathbf{u}$, the symbol decision thresholds for each of the D MZM output ports are set according to

$$\mathbf{I}_0 = \max_{\mathbf{u}} RG\mathbf{p}_0(\mathbf{u}), \quad \mathbf{I}_1 = \min_{\mathbf{u}} RG\mathbf{p}_1(\mathbf{u}), \tag{36}$$

$$\mathbf{I}_{\mathrm{dec}} = \frac{\mathbf{I}_0/\sigma_{n,0} + \mathbf{I}_1/\sigma_{n,1}}{\mathbf{1}/\sigma_{n,0} + \mathbf{1}/\sigma_{n,1}}. \tag{37}$$

The division operations in (37) are performed elementwise because $\mathbf{1}, \sigma_{n,0}$ and $\sigma_{n,1}$ are $D \times 1$ vectors. Then, using the symbol decision thresholds in (37), we can compute the $D \times 1$ vector of quality Q factors (one Q factor for each MZM port) as a function of $\mathbf{u}$ as

$$\mathbf{Q}_0(\mathbf{u}) = \left(\mathbf{I}_{\mathrm{dec}} - RG\mathbf{p}_0(\mathbf{u})\right)/\sigma_{\mathrm{n},0}(\mathbf{u}), \tag{38}$$

$$\mathbf{Q}_1(\mathbf{u}) = \left(RG\mathbf{p}_1(\mathbf{u}) - \mathbf{I}_{\mathrm{dec}}\right)/\sigma_{\mathrm{n},1}(\mathbf{u}), \tag{39}$$

$$\mathbf{Q}(\mathbf{u}) = \min\left(\mathbf{Q}_0(\mathbf{u}), \mathbf{Q}_1(\mathbf{u})\right). \tag{40}$$

The quality Q factor is based on the the tail probabilities of the standard normal distribution, and is commonly used to approximate BER in communications literature [35]. The average BER of the link across all D MZM ports can finally be computed by enumerating over all 4^D states of $\mathbf{u}$, given by

$$\mathrm{BER} = \frac{1}{D}\sum_{i=1}^{D}\sum_{\mathbf{u}} \frac{1}{2}\mathrm{erfc}\left(\frac{[\mathbf{Q}(\mathbf{u})]_i}{\sqrt{2}}\right) p(\mathbf{u}), \tag{41}$$

where $\mathrm{erfc}(\cdot)$ is the complementary error function.

REFERENCES

[1] D. A. B. Miller, "Self-configuring universal linear optical component," *Photon. Res.*, vol. 1, no. 1, pp. 1–15, Jun. 2013.

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

[3] T. Mizuno, H. Takara, K. Shibahara, A. Sano, and Y. Miyamoto, "Dense space division multiplexed transmission over multicore and multimode fiber for long-haul transport systems," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1484–1493, Mar. 2016.

[4] G. Rademacher *et al.*, "Long-haul transmission over few-mode fibers with space-division multiplexing," *J. Lightw. Technol.*, vol. 36, no. 6, pp. 1382–1388, Mar. 2018.

[5] H. Liu *et al.*, "Demonstration of stable 3×10 Gb/s mode group-multiplexed transmission over a 20 km few-mode fiber," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper W4J.2.

[6] I. Gasulla and J. M. Kahn, "Performance of direct-detection mode-group-division multiplexing using fused fiber couplers," *J. Lightw. Technol.*, vol. 33, no. 9, pp. 1748–1760, May 2015.

[7] B. Franz and H. Bülow, "Mode group division multiplexing in graded-index multimode fibers," *Bell Labs Tech. J.*, vol. 18, no. 3, pp. 153–172, Dec. 2013.

[8] J. Carpenter, B. C. Thomsen, and T. D. Wilkinson, "Degenerate mode-group division multiplexing," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3946–3952, Dec. 2012.

[9] 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.

[10] S. O. Arik and J. M. Kahn, "Low-complexity implementation of convex optimization-based phase retrieval," *J. Lightw. Technol.*, vol. 36, no. 12, pp. 2358–2365, Jun. 2018.
[11] A. Annoni *et al.*, "Unscrambling light—Automatically undoing strong mixing between modes," *Light: Sci. Appl.*, vol. 6, no. 12, 2017, Art. no. e17110.
[12] 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.
[13] 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.
[14] K. Choutagunta *et al.*, "Modal dynamics in spatially multiplexed links," in *Proc. Opt. Fiber Commun. Conf.*, 2019, Paper W4C.1.
[15] G. Labroille, B. Denolle, P. Jian, P. Genevaux, N. Treps, and J.-F. Morizur, "Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion," *Opt. Express*, vol. 22, no. 13, pp. 15 599–15 607, Jun. 2014.
[16] N. K. Fontaine *et al.*, "Packaged 45-mode multiplexers for a 50-μm graded index fiber," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2018, Paper M2I.2.
[17] 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, 2019, Art. no. 1865.
[18] R. Tang *et al.*, "Reconfigurable all-optical on-chip MIMO three-mode demultiplexing based on multi-plane light conversion," *Opt. Lett.*, vol. 43, no. 8, pp. 1798–1801, Apr. 2018.
[19] R. Tang *et al.*, "Reconfigurable 3-channel all-optical MIMO circuit on silicon based on multi-plane light conversion," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper W1E.3.
[20] 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.
[21] J. Zhou, J. Wu, and Q. Hu, "Tunable arbitrary unitary transformer based on multiple sections of multicore fibers with phase control," *Opt. Express*, vol. 26, no. 3, pp. 3020–3036, Feb. 2018.
[22] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Phys. Rev. Lett.*, vol. 73, pp. 58–61, Jul. 1994.
[23] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "Optimal design for universal multiport interferometers," *Optica*, vol. 3, no. 12, pp. 1460–1465, Dec. 2016.
[24] D. A. B. Miller, "Self-aligning universal beam coupler," *Opt. Express*, vol. 21, no. 5, pp. 6360–6370, Mar. 2013.
[25] F. Morichetti *et al.*, "Non-invasive on-chip light observation by contactless waveguide conductivity monitoring," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 4, pp. 292–301, Jul./Aug. 2014.
[26] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. J. Appl. Math.*, no. 2, pp. 164–168, 1944.
[27] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.
[28] J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," in *Numerical Analysis*, G. A. Watson, Ed. Berlin, Germany: Springer, 1978, pp. 105–116.
[29] J. J. Sylvester, "Thoughts on inverse orthogonal matrices, simultaneous sign successions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers," *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 34, no. 232, pp. 461–475, 1867.
[30] 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.
[31] P. Orlandi, F. Morichetti, M. J. Strain, M. Sorel, A. Melloni, and P. Bassi, "Tunable silicon photonics directional coupler driven by a transverse temperature gradient," *Opt. Lett.*, vol. 38, no. 6, pp. 863–865, Mar. 2013.
[32] 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.
[33] T. J. Seok, N. Quack, S. Han, W. Zhang, R. S. Muller, and M. C. Wu, "64×64 low-loss and broadband digital silicon photonic MEMS switches," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2015, pp. 1–3.
[34] J. K. Perin, A. Shastri, and J. M. Kahn, "Design of low-power DSP-free coherent receivers for data center links," *J. Lightw. Technol.*, vol. 35, no. 21, pp. 4650–4662, Nov. 2017.
[35] J. Proakis and M. Salehi, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2008.

Karthik Choutagunta (S'11) received the B.S. degree from the University of Texas at Austin, Austin, TX, USA, in 2014, and the M.S. degree in electrical engineering, in 2016, from Stanford University, Stanford, CA, USA, where he is currently working toward the Ph.D. degree. He was with Coherent Logix, Austin, TX, USA, in 2012, with Intel Corp., Austin, TX, USA, from 2012 to 2014, with Apple, Inc., Cupertino, CA, USA, in the fall of 2015 and the summer of 2016, with Facebook, Inc., Menlo Park, CA, USA, in 2017, and with Nokia Bell Laboratories, Crawford Hill Laboratory, Holmdel, NJ, USA, in 2018. His main current research interests include various topics in communications, signal processing, machine learning, and optimization.

Ian Roberts received the B.S. degree from the University of Ottawa, Ottawa, ON, Canada, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, USA, all in electrical engineering. He was with Ciena, working on DSP hardware development for WaveLogic 3 Extreme prior to beginning graduate studies. His research interests include optical communication, optimization, and compensation of nonlinear propagation. He is a Student Member of the OSA.

David A. B. Miller (M'83–F'95) received the Ph.D. degree in physics from Heriot-Watt University, Edinburgh, U.K., in 1979. He was with Bell Laboratories from 1981 to 1996, and as a Department Head after 1987. He is currently the W. M. Keck Professor of Electrical Engineering with Stanford University, Stanford, CA, USA. He was the President of the IEEE Lasers and Electro-Optics Society (now Photonics Society) in 1995. He has authored or coauthored more than 270 scientific papers and a textbook *Quantum Mechanics for Scientists and Engineers* (Cambridge Univ. Press, 2008), and holds 75 patents. His research interests include physics and devices in nanophotonics, nanometallics, and quantum well optoelectronics, and fundamentals and applications of optics in information sensing, switching, and processing. He received numerous awards. He is a Fellow of the Optical Society of America, the American Physical Society, the Royal Society, and the Royal Society of Edinburgh, and a Member of the National Academy of Sciences and the National Academy of Engineering.

Joseph M. Kahn (M'90–SM'98–F'00) received 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. In 1987–1990, he was with AT&T Bell Laboratories. In 1989, he demonstrated the first successful synchronous (i.e., coherent) detection in optical fiber systems, achieving record receiver sensitivity. In 1990–2003, he was at the Electrical Engineering and Computer Sciences Faculty, University of California, Berkeley. He demonstrated coherent detection of QPSK in 1992. In 1999, he and Shiu published the first work on shaping and nonequiprobable signaling for optical communications. In the 1990s and early 2000s, he and collaborators performed seminal work on indoor and outdoor free-space optical communications and multi-input multi-output wireless communications. In 2000, he and 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, he and Ho 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. He and collaborators have extensively studied rate-adaptive coding and modulation, as well as digital signal processing for mitigating linear and nonlinear impairments in coherent systems. In 2008, Kahn and Ip (and Li independently) invented simplified digital backpropagation for compensating fiber Kerr nonlinearity and dispersion. Since 2004, he 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 current research addresses 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 received the National Science Foundation Presidential Young Investigator Award in 1991. He is a Fellow of the OSA.