Optics Letters

Direct-detection mode-division multiplexing in modal basis using phase retrieval

Sercan Ö. Arik* and Joseph M. Kahn

Stanford University, E. L. Ginzton Laboratory, Department of Electrical Engineering, Stanford, California 94305, USA
**Corresponding author: soarik@stanford.edu*

Received 13 May 2016; revised 9 August 2016; accepted 10 August 2016; posted 10 August 2016 (Doc. ID 265153); published 9 September 2016

Mode-division multiplexing (MDM) can increase the capacity of direct-detection short-reach systems in proportion to the number of modes employed. MDM requires compensation of modal crosstalk at a transmitter or receiver by the multi-input multi-output (MIMO) signal processing. We show that the channel estimation required for the MIMO processing in a basis of modes can be expressed as a phase retrieval problem. We propose three techniques for the estimation: sparse training sequences, convex optimization (CO) and alternating minimization. We demonstrate the superior performance of the CO technique.

OCIS codes: (060.2330) Fiber optics communications; (060.4230) Multiplexing; (070.1170) Analog optical signal processing.

http://dx.doi.org/10.1364/OL.41.004265

Multi-mode fibers (MMFs) traditionally have been used in short-reach links without exploiting the plurality of modes [1]. In an MMF supporting D orthogonal propagating spatial-polarization modes, the information throughput can be increased up to D-fold by multiplexing in the modal degrees of freedom [2]. A simple, but suboptimal, approach is mode-group-division multiplexing (MGDM), which maps/demaps data streams to/from groups of modes having nearly degenerate propagation constants. MGDM depends on minimizing coupling between the mode groups by careful link design [3,4]. In MGDM, the number of mode groups available for multiplexing is much less than the number of modes, scaling roughly as $\sqrt{D}$ for large D. A throughput increase proportional to D can be obtained using mode-division multiplexing (MDM), which maps/demaps data streams to/from individual propagating modes. A major challenge to MDM is posed by coupling between modes, which causes crosstalk between multiplexed data streams. Mode coupling within a group (intra-group coupling) is typically much stronger than that between different groups (inter-group coupling) [5].

MDM systems typically compensate for crosstalk using adaptive channel estimation and multi-input multi-output (MIMO) signal processing. In long-haul MDM systems using coherent detection, both the intensities and phases of received modal fields are available for the estimation [6]. In short-reach systems, direct detection (DD) is preferred for its low cost. Since only the intensities of the received modal fields are available in DD systems, the MIMO channel estimation methods used in coherent systems are not applicable. Prior work on DD-MDM has assumed that multiplexed data streams are mapped/demapped to/from spatially non-overlapping regions of the multi-mode beam, and the MIMO signal processing is performed in the basis of spatial samples [7,8]. This approach does not enable the MIMO signal processing to be performed separately for different mode groups, which may yield good performance while greatly reducing complexity in the systems having negligible inter-group coupling. In this Letter, we study DD-MDM link architectures and MIMO estimation techniques using a modal basis, which enable separate processing of different mode groups, and may yield throughput approaching DD capacity limits.

Three canonical link architectures for DD-MDM in a modal basis are shown in Figs. 1(a)–1(c). In all cases, D data streams are mapped to/from D propagating spatial-polarization modes using optical multiplexer/demultiplexer devices, which can be implemented with low loss and high modal selectivity

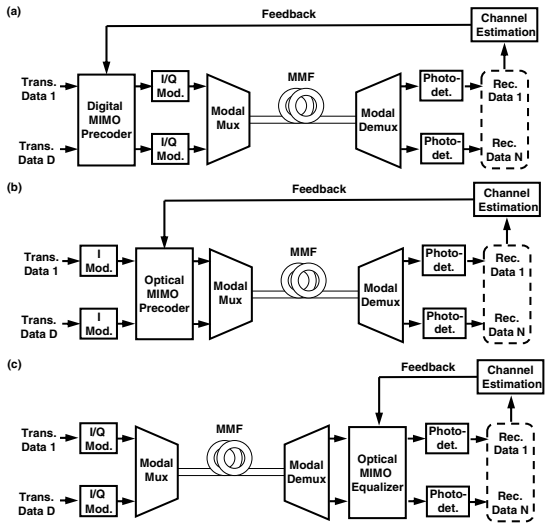

Fig. 1. DD-MDM system with (a) digital MIMO precoding at the transmitter, (b) optical MIMO precoding at the transmitter, and (c) optical MIMO equalization at the receiver.

[9]. The three schemes differ in the placement and method of the MIMO signal processing used to compensate for modal crosstalk. Digital MIMO precoding, shown in Fig. 1(a), requires high-speed digital signal processing and digital-to-analog conversion, and a dual-quadrature modulator for each mode. Optical MIMO precoding and equalization, shown in Figs. 1(b) and 1(c), respectively, may be performed without fundamental loss using arrays of Mach–Zehnder interferometers [10] or micro-ring resonators [11]. Although intensity modulation is employed, dual-quadrature modulators are required in Fig. 1(c) to generate complex-valued training sequences for some estimation methods. In Fig. 1(b), single-quadrature modulators are sufficient, as complex-valued training sequences can be obtained with the aid of the MIMO precoder, which is not otherwise used during training. All three options in Figs. 1(a)–1(c) require high-speed analog-to-digital conversion and computation for the channel estimation.

We assume the link is sufficiently short so that the frequency selectivity of the modal dispersion is negligible. (The generalization to frequency-selective channels may employ frequency-selective digital or optical MIMO processing.) The transmitted and received baseband modal field envelopes for an nth sequence are given by

$$\mathbf{y}^{(n)} = \mathbf{M}^{(n)}\mathbf{H}\mathbf{P}^{(n)}\mathbf{x}^{(n)}, \tag{1}$$

where $\mathbf{x}^{(n)}$ is a $D \times 1$ transmitted vector, $\mathbf{P}^{(n)}$ is a $D \times D$ transmitter MIMO precoder matrix, $\mathbf{H}$ is a $D \times D$ channel transfer matrix, $\mathbf{M}^{(n)}$ is a $D \times D$ receiver MIMO equalizer matrix, and $\mathbf{y}^{(n)}$ is a $D \times 1$ received vector. Assuming a thermal noise-limited DD receiver, a photocurrent proportional to $\mathbf{d}_i^{(n)} = |\mathbf{y}_i^{(n)}|^2 + \mathbf{n}_i^{(n)}$ is obtained for each mode, $1 \le i \le D$. The additive noise $\mathbf{n}^{(n)}$ is temporally white and has a multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance $\mathbf{I}\sigma^2$, where $\mathbf{I}$ is a $D \times D$ identity matrix. The system electrical signal-to-noise ratio (SNR) is defined as

$$\mathrm{SNR} = \left[\left\langle \sum_{i=1}^{D} |\mathbf{y}_i^{(n)}|^2 \right\rangle / (D \cdot \sigma)\right]^2, \tag{2}$$

where $\langle\,\rangle$ denotes averaging over training sequences. Since (2) is proportional to the square of the optical power per mode, a 1 dB change of the optical power per mode corresponds to a 2 dB change of $10\log_{10}(\mathrm{SNR})$.

We focus first on systems using transmitter MIMO precoding, as shown in Figs. 1(a) and 1(b). Initially, we choose $\mathbf{P}^{(n)} = \mathbf{I}$ and $\mathbf{M}^{(n)} = \mathbf{I}$, and we estimate $\mathbf{P}^{(n)}$ using training sequences. The estimation of $\mathbf{M}^{(n)}$ is not possible here. Since DD causes loss of a phase common to all modes, the best estimate of the channel is

$$\tilde{\mathbf{H}} = \mathbf{D} \cdot \mathbf{H}, \tag{3}$$

where $\mathbf{D} = \mathrm{diag}(\boldsymbol{\varsigma})$ is a diagonal unitary matrix. To fully compensate for crosstalk, a MIMO equalizer must yield $\mathbf{M}^{(n)} \cdot \mathbf{H} = \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a diagonal matrix. For the best case (3), $\mathbf{M}^{(n)} = \boldsymbol{\Gamma} \cdot \tilde{\mathbf{H}}^{-1}\mathbf{D}^{-1}$. To determine $\mathbf{M}^{(n)}$, there must be a unique decomposition of $\tilde{\mathbf{H}}$ in the form $\boldsymbol{\Lambda} \cdot \mathbf{B}$, where $\boldsymbol{\Lambda}$ is a diagonal unitary matrix, $\boldsymbol{\Lambda} \cdot \mathbf{D} = \mathbf{I}$, and $\mathbf{B}$ is an arbitrary matrix. However, there are infinitely many decompositions of $\tilde{\mathbf{H}}$ in the form $\boldsymbol{\Lambda} \cdot \mathbf{B}$ and $\boldsymbol{\Lambda} \cdot \mathbf{D} \ne \mathbf{I}$ for all, except for $\boldsymbol{\Lambda} = \mathbf{D}^H$. Hence, $\mathbf{M}^{(n)}$ cannot be identified. On the other hand, as a MIMO precoder, ideally $\tilde{\mathbf{H}}^{-1}$ can be used to fully compensate for crosstalk because $\mathbf{H} \cdot \tilde{\mathbf{H}}^{-1} = \mathbf{D}^{-1}$, which is a diagonal unitary matrix. The estimation of $\tilde{\mathbf{H}}$ involves phase retrieval, as the phase information is extracted from the measured intensities of the inner products (rows of $\tilde{\mathbf{H}}$) using known vectors (training sequences $\mathbf{x}^{(n)}$).

Second, we focus on systems using receiver MIMO equalization, as shown in Fig. 1(c). While estimating the MIMO equalizer $\mathbf{M}^{(n)}$, the knowledge of $\tilde{\mathbf{H}}$ is used and, hence, the estimation of the MIMO precoder is a prerequisite. In addition, the application of known MIMO equalizer masks is required. We showed above that when $\mathbf{M}^{(n)} = \mathbf{I}$, the ambiguity of the factor $\mathbf{D} = \mathrm{diag}(\boldsymbol{\varsigma})$ in (3) cannot be avoided. By using MIMO equalizer masks $\mathbf{M}^{(n)} \ne \mathbf{I}$ (which are not necessarily related to channel estimates), modal interferometry is performed, and $\mathbf{D} = \mathrm{diag}(\boldsymbol{\varsigma})$ can be estimated up to an overall scalar phase factor. As a MIMO equalizer, ideally $\tilde{\mathbf{H}}^{-1} \cdot \mathbf{D}$ can be used to fully compensate for crosstalk, since $\tilde{\mathbf{H}}^{-1} \cdot \mathbf{D} \cdot \mathbf{H} = \mathbf{I}$. For the estimation of $\boldsymbol{\varsigma}$, we observe that $\mathbf{y}_d^{(n)}$ in (1) can be expressed, for each $1 \le d \le D$, as

$$\mathbf{y}_d^{(n)} = \sum_{i=1}^{D} \boldsymbol{\psi}_{di}^{(n)} \cdot \boldsymbol{\varsigma}_i^*, \tag{4}$$

where $\boldsymbol{\psi}_{di}^{(n)} = \mathbf{M}_{di}^{(n)} \cdot \boldsymbol{\alpha}_i^{(n)}$ and $\boldsymbol{\alpha}^{(n)} = \tilde{\mathbf{H}}\mathbf{x}^{(n)}$. Given $\mathbf{x}^{(n)}$, $\tilde{\mathbf{H}}$, and $\mathbf{M}^{(n)}$, the estimation of $\boldsymbol{\varsigma}$ involves phase retrieval, as the phase information is extracted from measured intensities of the inner products of $\boldsymbol{\varsigma}$ with known vectors $\boldsymbol{\psi}_d^{(n)*}$.

We now present three techniques to estimate $\tilde{\mathbf{H}}$ and/or $\boldsymbol{\varsigma}$ and to enable transmitter or receiver MIMO processing.

(1) *Sparse training sequences* (*STSs*): STSs yield the sequential estimation of the phases and intensities through their sparse structures [12]. Each batch of D^2 sequences contains the following three subsets [12]:

$$\underset{1 \le i \le D}{\mathbf{S}_{1,ia}} = \delta_{a,i}, \tag{5}$$

$$\underset{1 \le i \le D, i < k < D}{\mathbf{S}_{2,ika}} = \delta_{a,i}/\sqrt{2} + \delta_{a,k}/\sqrt{2}, \tag{6}$$

$$\underset{1 \le i \le D, i < k < D}{\mathbf{S}_{3,ika}} = \delta_{a,i}/\sqrt{2} + j\delta_{a,k}/\sqrt{2}, \tag{7}$$

where δ is the Kronecker delta.

To estimate $\tilde{\mathbf{H}}$, batches of STSs are transmitted sequentially. For the sth batch (where $(s-1)D^2 + 1 \le n \le sD^2$), sequences (5)–(7) yield the measurements

$$\mathbf{m}_{1,ia}^{(s)} = |\mathbf{H}_{ai}|^2 + \mathbf{n}_{1,ia}, \tag{8}$$

$$\mathbf{m}_{2,ika}^{(s)} = 0.5(|\mathbf{H}_{ai}|^2 + |\mathbf{H}_{ak}|^2) + |\mathbf{H}_{ai}||\mathbf{H}_{ak}|\cos(\varphi(\mathbf{H}_{ai}) - \varphi(\mathbf{H}_{ak})) + \mathbf{n}_{2,ika}, \tag{9}$$

$$\mathbf{m}_{3,ika}^{(s)} = 0.5(|\mathbf{H}_{ai}|^2 + |\mathbf{H}_{ak}|^2) + |\mathbf{H}_{ai}||\mathbf{H}_{ak}|\sin(\varphi(\mathbf{H}_{ai}) - \varphi(\mathbf{H}_{ak})) + \mathbf{n}_{3,ika}, \tag{10}$$

respectively, where $\varphi(\,)$ is the argument function. The maximum likelihood (ML) estimation of $\tilde{\mathbf{H}}$ given (8)–(10) does not have a closed-form solution and is numerically complex. To reduce complexity, the intensity and phase estimations are separated, which is the principal advantage of the STSs (5)–(7), in contrast to other possible training sequences. For the intensities, the ML estimates only using $\mathbf{m}_{1,ia}^{(s)}$ are used:

$$|\tilde{\mathbf{H}}_{ik}^{(s)}| \leftarrow \sqrt{\mathbf{m}_{1,ik}}. \tag{11}$$

For the phases, $\boldsymbol{\gamma}_{ika}^{(s)} = \mathbf{m}_{2,ika}^{(s)} + j\mathbf{m}_{3,ika}^{(s)} - (0.5 + 0.5j)(\mathbf{m}_{1,ia}^{(s)} + \mathbf{m}_{1,ka}^{(s)})$ are used. Ideally, $\varphi(\widetilde{\mathbf{H}}_{ai}^{(s)}) - \varphi(\widetilde{\mathbf{H}}_{ak}^{(s)}) = \varphi(\boldsymbol{\gamma}_{ika}^{(s)})$. The estimation is based on imposing this constraint sequentially, which corresponds to sequentially finding ML estimates for the phases only using $\mathbf{m}_{2,ika}^{(s)}$ and $\mathbf{m}_{3,ika}^{(s)}$. Since the phase common to all modes is lost, we set $\varphi(\widetilde{\mathbf{H}}_{ii}^{(s)}) \leftarrow 0$ and estimate the phases of non-diagonal elements with respect to diagonal elements as

$$\varphi(\widetilde{\mathbf{H}}_{ik}^{(s)}) \leftarrow -\varphi(\boldsymbol{\gamma}_{iki}^{(s)}), \tag{12}$$

$$\varphi(\widetilde{\mathbf{H}}_{ki}^{(s)}) \leftarrow \varphi(\boldsymbol{\gamma}_{ikk}^{(s)}). \tag{13}$$

To incorporate the differential phase information between the non-diagonal terms, we further impose the following sequentially:

$$\varphi(\widetilde{\mathbf{H}}_{ai}^{(s)}) \leftarrow \varphi(\widetilde{\mathbf{H}}_{ai}^{(s)}) - 0.5(\varphi(\widetilde{\mathbf{H}}_{ai}^{(s)}) - \varphi(\widetilde{\mathbf{H}}_{ak}^{(s)}) - \varphi(\boldsymbol{\gamma}_{ika}^{(s)})), \tag{14}$$

$$\varphi(\widetilde{\mathbf{H}}_{ak}^{(s)}) \leftarrow \varphi(\widetilde{\mathbf{H}}_{ak}^{(s)}) + 0.5(\varphi(\widetilde{\mathbf{H}}_{ai}^{(s)}) - \varphi(\widetilde{\mathbf{H}}_{ak}^{(s)}) - \varphi(\boldsymbol{\gamma}_{ika}^{(s)})). \tag{15}$$

Eventually, the estimates obtained using (11)–(15) for the sth batch are incorporated into the overall estimates as

$$|\widetilde{\mathbf{H}}_{ik}| \leftarrow \sqrt{(1/s)|\widetilde{\mathbf{H}}_{ik}^{(s)}|^2 + ((s-1)/s)|\widetilde{\mathbf{H}}_{ik}|^2}, \tag{16}$$

$$\varphi(\widetilde{\mathbf{H}}_{ik}) \leftarrow \varphi((1/s)\exp(j\widetilde{\mathbf{H}}_{ik}^{(s)}) + ((s-1)/s)\exp(j\widetilde{\mathbf{H}}_{ik})). \tag{17}$$

To estimate $\boldsymbol{\varsigma}$, STSs are used as the rows of the MIMO equalizer mask $\mathbf{M}^{(n)}$. They yield the measurements

$$\mathbf{d}_{1,i}^{(n)} = |\boldsymbol{\alpha}_i^{(n)}|^2 + \mathbf{n}_{1,i}, \tag{18}$$

$$\mathbf{d}_{2,ik}^{(n)} = 0.5(|\boldsymbol{\alpha}_i^{(n)}|^2 + |\boldsymbol{\alpha}_k^{(n)}|^2) + |\boldsymbol{\alpha}_i^{(n)}||\boldsymbol{\alpha}_k^{(n)}|\cos(\varphi(\boldsymbol{\varsigma}_i) - \varphi(\boldsymbol{\varsigma}_k) + \varphi(\boldsymbol{\alpha}_i^{(n)}) - \varphi(\boldsymbol{\alpha}_k^{(n)})) + \mathbf{n}_{2,i}, \tag{19}$$

$$\mathbf{d}_{3,ik}^{(n)} = 0.5(|\boldsymbol{\alpha}_i^{(n)}|^2 + |\boldsymbol{\alpha}_k^{(n)}|^2) + |\boldsymbol{\alpha}_i^{(n)}||\boldsymbol{\alpha}_k^{(n)}|\sin(\varphi(\boldsymbol{\varsigma}_i) - \varphi(\boldsymbol{\varsigma}_k) + \varphi(\boldsymbol{\alpha}_i^{(n)}) - \varphi(\boldsymbol{\alpha}_k^{(n)})) + \mathbf{n}_{3,i}. \tag{20}$$

Unlike the estimation of $\widetilde{\mathbf{H}}$, the intensity estimation is not required since $|\varsigma_i| = 1$. To estimate the differential phase terms $\varphi(\boldsymbol{\varsigma}_i) - \varphi(\boldsymbol{\varsigma}_k)$, $\boldsymbol{\eta}_{ik}^{(n)} = \mathbf{d}_{2,ik}^{(n)} + j\mathbf{d}_{3,ik}^{(n)} - (0.5 + j0.5)(\mathbf{d}_{1,i}^{(n)} + \mathbf{d}_{1,k}^{(n)})$ are used. To construct each $\boldsymbol{\eta}_{ik}^{(n)}$, four measurements are needed. Each training sequence can yield $\lfloor D/4 \rfloor$ terms, where $\lfloor\ \rfloor$ is the floor function. (When $D < 4$, splitting of the output is required to provide measurement diversity.) In total, there are $(D^2 - D)/2$ possible $\boldsymbol{\eta}_{ik}^{(n)}$ terms. A batch estimation is applied after obtaining all $\boldsymbol{\eta}_{ik}^{(n)}$ terms. The batch size is $L_B = \lceil (D^2 - D)/(2(\lfloor D/4 \rfloor)) \rceil$ and, for the sth batch, $(s-1)L_B + 1 \le n \le sL_B$. Since the phase common to all modes is lost, we set $\varphi(\boldsymbol{\varsigma}_1^{(s)}) = 0$ and, similar to (12),(13), obtain the phase estimates

$$\varphi(\hat{\boldsymbol{\varsigma}}_k^{(s)}) \leftarrow \varphi(\boldsymbol{\alpha}_1^{(n)}) - \varphi(\boldsymbol{\alpha}_k^{(n)}) - \varphi(\boldsymbol{\eta}_{1k}^{(n)}). \tag{21}$$

Similar to (14) and (15), the information for relative phase differences is imposed sequentially:

$$\varphi(\hat{\boldsymbol{\varsigma}}_i^{(s)}) \leftarrow \varphi(\hat{\boldsymbol{\varsigma}}_i^{(s)}) - 0.5(\varphi(\boldsymbol{\alpha}_i^{(n)}) - \varphi(\boldsymbol{\alpha}_k^{(n)}) + \varphi(\hat{\boldsymbol{\varsigma}}_i^{(s)}) - \varphi(\hat{\boldsymbol{\varsigma}}_k^{(s)}) - \varphi(\eta_{ik}^{(n)})), \tag{22}$$

$$\varphi(\hat{\boldsymbol{\varsigma}}_k^{(s)}) \leftarrow \varphi(\hat{\boldsymbol{\varsigma}}_k^{(s)}) + 0.5(\varphi(\boldsymbol{\alpha}_i^{(n)}) - \varphi(\boldsymbol{\alpha}_k^{(n)}) + \varphi(\hat{\boldsymbol{\varsigma}}_i^{(s)}) - \varphi(\hat{\boldsymbol{\varsigma}}_k^{(s)}) - \varphi(\eta_{ik}^{(n)})). \tag{23}$$

Lastly, similar to (17), the estimate after the sth batch is incorporated into the overall estimate as

$$\varphi(\hat{\boldsymbol{\varsigma}}) \leftarrow (1/s)\varphi(\hat{\boldsymbol{\varsigma}}^{(s)}) + ((s-1)/s)\varphi(\hat{\boldsymbol{\varsigma}}). \tag{24}$$

A fundamental limitation of STSs is that only a small fraction of the training sequences and equalizer masks are used for the estimation of a small fraction of the unknowns. The following two techniques are considered to overcome this limitation.

(2) *Convex optimization* (*CO*): an efficient strategy for phase retrieval is to represent the unknowns in a higher dimensional space, where efficient optimization techniques can be applied [13].

To estimate $\widetilde{\mathbf{H}}$, a key observation is that the noiseless channel is equivalent to the following problem, for each $1 \le i \le D$:

$$\min \operatorname{rank}(\mathbf{C}_i) \ \text{s.t.}\ \mathbf{C}_i \succeq 0 \quad \text{and} \quad \mathbf{x}^{(n)H}\mathbf{C}_i\mathbf{x}^{(n)} = \mathbf{d}_i^{(n)}. \tag{25}$$

The solution of (25) is 1 for the optimal matrix parameter $\hat{\mathbf{C}}_i = \mathbf{h}_i^H\mathbf{h}_i$, where $\mathbf{h}_i$ is the ith row of $\mathbf{H}$. Rank-minimization problems are computationally non-deterministic polynomial-time hard. A common convex relaxation technique is to replace the rank function by the trace function, which provides an approximate solution as a semi-definite program [13]. In the presence of noise, the constraint $\mathbf{x}^{(n)H}\mathbf{C}_i\mathbf{x}^{(n)} = \mathbf{d}_i^{(n)}$ can be numerically infeasible, especially for low SNR. Instead, we consider an ℓ^2 norm penalty term, which also corresponds to the ML estimation for spatially white and zero-mean additive Gaussian noise. Overall, (25) becomes the following convex program:

$$\min\ \operatorname{tr}(\mathbf{C}_i) + \lambda\sum\nolimits_{n=1}^{N}|\mathbf{x}^{(n)H}\mathbf{C}_i\mathbf{x}^{(n)} - \mathbf{d}_i^{(n)}|^2 \quad \text{s.t.}\ \mathbf{C}_i \succeq 0, \tag{26}$$

where λ is a trade-off constant. For the optimal matrix parameter $\hat{\mathbf{C}}_i$, the rank-1 component yields an estimate of the ith row of the channel matrix. Given the principal component analyses $\hat{\mathbf{C}}_i = \sum_{k=1}^{D}\hat{\chi}_{i,k}\hat{\mathbf{u}}_{i,k}\hat{\mathbf{u}}_{i,k}^H$, we obtain

$$\widetilde{\mathbf{H}} \leftarrow \left[\sqrt{\hat{\chi}_{1,1}}\hat{\mathbf{u}}_{1,1}^T \quad \cdots \quad \sqrt{\hat{\chi}_{D,1}}\hat{\mathbf{u}}_{D,1}^T\right]^T. \tag{27}$$

Note that, unlike the STS, there are no restrictions on the training sequences; they can be generated randomly.

For the estimation of $\boldsymbol{\varsigma}$, we adapt (26) and obtain

$$\min \operatorname{tr}(\mathbf{K}) + \lambda\sum\nolimits_{n}\sum\nolimits_{d=1}^{D}|\boldsymbol{\psi}_d^{(n)H}\mathbf{K}\boldsymbol{\psi}_d^{(n)} - \mathbf{d}_d^{(n)}|^2 \ \text{s.t.}\ \mathbf{K} \succeq 0. \tag{28}$$

For the optimal matrix parameter $\hat{\mathbf{K}} = \sum_{k=1}^{D}\hat{\chi}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^H$, similar to (27), the estimate is given by

$$\hat{\boldsymbol{\varsigma}} \leftarrow \sqrt{\hat{\chi}_1}(\hat{\mathbf{u}}_1)^*. \tag{29}$$

Since $\boldsymbol{\varsigma}$ has elements with unit magnitude, the estimation can be further improved by debiasing as $\hat{\boldsymbol{\varsigma}}_i \leftarrow \exp(j\varphi(\hat{\boldsymbol{\varsigma}}_i))$. Note that, unlike STS, there is no restriction on the known MIMO equalizer masks; they can be random.

(3) *Alternating minimization* (*AM*): an efficient optimization procedure is alternating the estimation of the missing phase information and the candidate solutions [14].

To estimate $\widetilde{\mathbf{H}}$, the alternating update equations for two unknown variables, the measured phases and the channel transfer matrix rows are

$$\mathbf{p}_i^{(n)} \leftarrow \exp(j\varphi(\mathbf{x}^{(n)H}\mathbf{v}_i)), \tag{30}$$

$$\mathbf{v}_i \leftarrow \arg\min_{\mathbf{v}} \left(\sum_{n=1}^{N} |\mathbf{x}^{(n)H}\mathbf{v} - \mathbf{p}_i^{(n)} \sqrt{\mathbf{d}_i^{(n)}}|^2 \right). \tag{31}$$

Step (31) is simply a least-squares problem that admits an analytical solution. After a sufficient number of iterations, $\mathbf{v}_i$ converges to an estimate of the ith row of $\widetilde{\mathbf{H}}$. To ensure reliable and fast convergence, choosing a good initial point of $\mathbf{v}_i$ is crucial. We use the first principal component of $1/N \sum_{n=1}^{N} \mathbf{d}_i^{(n)} \mathbf{x}^{(n)} \mathbf{x}^{(n)H}$ as the initial point, which guarantees convergence under certain assumptions [14]. As in the CO technique, there are no restrictions on the training sequences.

To estimate $\boldsymbol{\varsigma}$, the update equations are

$$\mathbf{q}_d^{(n)} \leftarrow \exp(j\varphi(\boldsymbol{\psi}_d^{(n)H}\mathbf{w})), \tag{32}$$

$$\mathbf{w} \leftarrow \arg\min_{\mathbf{v}} \left(\sum_{n} \sum_{d=1}^{D} |\boldsymbol{\psi}_d^{(n)H}\mathbf{w} - \mathbf{q}_d^{(n)} \sqrt{\mathbf{d}_d^{(n)}}|^2 \right). \tag{33}$$

After a sufficient number of iterations, $\mathbf{w}^*$ converges to an estimate of $\boldsymbol{\varsigma}$. Similar to CO, debiasing is applied to improve accuracy. The initial point used is the first principal component of $1/D\langle \sum_{d=1}^{D} \mathbf{d}_d^{(n)} \boldsymbol{\psi}_d^{(n)} \boldsymbol{\psi}_d^{(n)H} \rangle$. As in the CO technique, there are no restrictions on the known MIMO equalizer masks.

We demonstrate the performance of the estimation techniques for a DD-MDM system employing on-off keying. We generate fully random $D \times D$ unitary channel matrices (describing worst-case full intra- and inter-group coupling) and use the estimation techniques to enable MIMO processing at the transmitter or receiver. We assume that the same technique is used to estimate $\widetilde{\mathbf{H}}$ and $\boldsymbol{\varsigma}$ for simplicity, although, in principle, a combination of different techniques is possible. As a reference, a back-to-back (B2B) system ($\mathbf{M}^{(n)}\mathbf{H}\mathbf{P}^{(n)} = \mathbf{I}$) is considered. Figures 2(a) and 2(b) show the average bit-error ratio (BER) versus SNR for the estimation after 300 sequences. We observe that CO yields excellent performance in both cases, incurring SNR penalties less than 0.5 and 1 dB for the transmitter and receiver MIMO, respectively, assuming realistic target BERs. AM yields very good performance for the transmitter MIMO, but yields a substantial penalty for the receiver MIMO. STS yields inferior performance for both cases, resulting in substantial SNR penalties. The receiver MIMO yields higher BERs for all techniques because the error in the estimation of $\widetilde{\mathbf{H}}$ propagates in the estimation of $\boldsymbol{\varsigma}$.

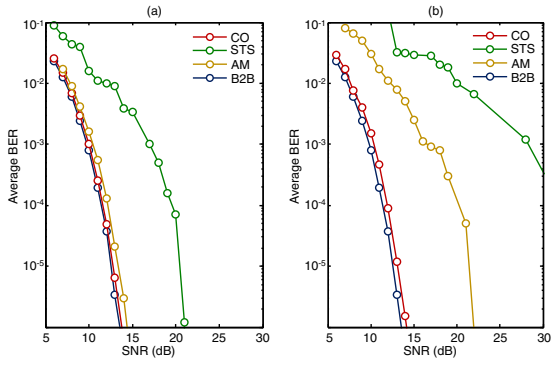

Fig. 2. Average BER versus SNR for (a) transmitter MIMO processing and (b) receiver MIMO processing. Assuming $D = 6$, $N = 300$, and $\lambda = 10$, Monte Carlo simulations are averaged over 20 channel realizations and 6,400,000 transmitted sequences. CO problems are solved using [15].

Given typical short-reach link environments (e.g., data centers) and lengths (e.g., 0.1–1 km), the time scale for channel variations is expected to be of the order 0.1–10 ms. For $D \sim 2$–20 (requiring $N \sim 100$–1000) and practical symbol rates (~30–100 Gbaud), the timescale for transmission of the training sequences is of the order 1–30 ns. Hence, if the estimation techniques can be implemented in embedded hardware with a negligible computational delay, the overhead (loss of throughput) will be negligible. Even if power consumption is high during the estimation, the average power consumption per information symbol is expected to be low. Our initial studies suggest that sufficiently fast and low-complexity implementation of the techniques is possible, and these topics will be addressed in a future paper.

While MDM system capacity ideally scales in proportion to D, the estimation time and power consumption may scale faster, particularly in systems with full intra- and inter-group coupling. Scaling to large D can be facilitated by a judicious link design in to make inter-group coupling negligible [3–5], so that the channel matrix $\mathbf{H}$ becomes sparse, i.e., block diagonal. In this scenario, the MIMO methods described here can be applied to each mode group separately, dramatically reducing the estimation time and power consumption.

In conclusion, we described DD-MDM systems using digital or optical MIMO precoding or optical MIMO equalization, and three phase retrieval-based MIMO channel estimation methods. Among the three methods, CO yields the best performance, which is nearly ideal. Hardware-specific implementation of these methods, and the performance impact of mode-dependent loss, channel dynamics, and other effects are important topics for future study.

Funding. Huawei Technologies Co., Ltd.

REFERENCES

1. R. E. Freund, C. A. Bunge, N. N. Ledentsov, D. Molin, and C. Caspar, in *Proceeding of Optical Fiber Communication Conference*, San Diego, California, USA, March 2009, paper OMS1.
2. S. O. Arik and J. M. Kahn, Opt. Lett. **39**, 3258 (2014).
3. I. Gasulla and J. M. Kahn, J. Lightwave Technol. **33**, 1748 (2015).
4. G. Labroille, P. Jian, L. Garcia, J. B. Trinel, R. Kassi, L. Bigot, and J. F. Morizur, in *European Conference on Optical Communication (ECOC)*, Valencia, Spain, 2015, paper P.5.12.
5. K.-P. Ho and J. M. Kahn, in *Optical Fiber Telecommunications VI* (Elsevier, 2013), pp. 491.
6. S. O. Arik, J. M. Kahn, and K.-P. Ho, IEEE Signal Process. Mag. **31**(2), 25 (2014).
7. M. Nazarathy and A. Agmon, J. Lightwave Technol. **26**, 2037 (2008).
8. A. Spalvieri, P. Boffi, S. Pecorino, L. Barletta, M. Magarini, A. Gatto, P. Martelli, and M. Martinelli, Opt. Express **21**, 25174 (2013).
9. G. Labroille, P. Jian, N. Barré, B. Denolle, and J. F. Morizur, in *Optical Fiber Communication Conference*, Anaheim, California, USA, March 2016, paper Th3E.5.
10. D. A. B. Miller, Opt. Express **21**, 6360 (2013).
11. L. Yang, R. Ji, L. Zhang, J. Ding, and Q. Xu, Opt. Express **20**, 13560 (2012).
12. A. Agmon and M. Nazarathy, Opt. Express **15**, 13123 (2007).
13. E. J. Candes, T. Strohmer, and V. Voroninski, Commun. Pure Appl. Math. **66**, 1241 (2013).
14. P. Netrapalli, P. Jain, and S. Sanghavi, IEEE Trans. Signal Process. **63**, 4814 (2015).
15. M. Grant and S. Boyd, http://cvxr.com/cvx.