Adaptive Frequency-Domain Equalization in Mode-Division Multiplexing Systems

Sercan Ö. Arık, Daulet Askarov, and Joseph M. Kahn, *Fellow, IEEE*

***Abstract*—Long-haul mode-division multiplexing (MDM) employs adaptive multi-input multi-output (MIMO) equalization to compensate for modal crosstalk and modal dispersion. MDM systems must typically use MIMO frequency-domain equalization (FDE) to minimize computational complexity, in contrast to polarization-division-multiplexed systems in single-mode fiber, where time-domain equalization (TDE) has low complexity and is often employed to compensate for polarization effects. We study two adaptive algorithms for MIMO FDE: least mean squares (LMS) and recursive least squares (RLS). We analyze tradeoffs between computational complexity, cyclic prefix efficiency, adaptation time and output symbol-error ratio (SER), and the impact of channel group delay spread and fast Fourier transform (FFT) block length on these. Using FDE, computational complexity increases sublinearly with the number of modes, in contrast to TDE. Adaptation to an initially unknown fiber can be achieved in ∼3–5 μs using RLS or ∼15–25 μs using LMS in fibers supporting 6–30 modes. As compared to LMS, RLS achieves faster adaptation, higher cyclic prefix efficiency, lower SER, and greater tolerance to mode-dependent loss, but at the cost of higher complexity per FFT block. To ensure low computational complexity and fast adaptation in an MDM system, a low overall group delay spread is required. This is achieved here by a family of graded-index graded depressed-cladding fibers in which the uncoupled group delay spread decreases with an increasing number of modes, in concert with strong mode coupling.**

***Index Terms*—DSP complexity, MIMO, equalization, few-mode fiber, modal dispersion, mode coupling, mode-division multiplexing, multi-mode coherent receiver, multi-mode fiber, receiver signal processing.**

I. Introduction

THE continued exponential growth of data traffic has motivated research on increasing capacity in long-haul transmission systems. Information-theoretic limits of single-mode fiber (SMF) transmission imposed by noise, fiber nonlinearity and dispersion are being approached [1], [2] and compensation of these combined effects requires very high computational complexity [3], [4]. Increasing per-fiber capacity can be achieved more readily by increasing spatial dimensionality using multi-core fiber [5], [6] or multi-mode fiber (MMF) [5], [7] with multi-input multi-output (MIMO) transmission [8], [9]. Throughout this paper, D denotes the total number of dimensions available for multiplexing, including spatial and polarization degrees of freedom.

Long-haul transmission systems already employ multiplexing in the two polarization modes of SMF ($D = 2$). This is enabled by coherent detection and digital equalization [10], [11], which can compensate for chromatic dispersion (CD) and polarization-mode dispersion (PMD). Most such systems employ a hybrid equalization approach. CD, which is essentially fixed but has a long impulse response duration, is compensated by programmable (but not adaptive) frequency-domain equalization (FDE). By contrast, PMD, which can vary on a microsecond time scale [12], [13] but has a very short impulse response duration, is often compensated by adaptive 2×2 MIMO time-domain equalization (TDE) [14]–[18]. Filter taps are updated using either blind or data-aided methods such as the least mean squares (LMS) algorithm, typically using training sequences for initial adaptation and decision-directed updates afterwards [10], [17].

In systems using mode-division multiplexing (MDM) in MMFs ($D > 2$), receiver computational complexity increases because of an increase in D and because of the large group delay (GD) spread from modal dispersion (MD). Two approaches for minimizing GD spread and controlling receiver complexity are optimization of the fiber index profile and introduction of strong mode coupling [19]. For two mode groups ($D = 6$), fibers with very low uncoupled GD spread can be realized by choosing a core radius at which the GD-versus-radius curves for the two mode groups intersect [20]. For more than two mode groups ($D > 6$), this approach is not viable, since the curves for different pairs of modes intersect at different radii, and step-index fibers have prohibitively high GD spreads [21]. On the other hand, graded-index fibers with large cores ($D \to \infty$) have very low GD spreads [22], but for small D, graded-index fibers have prohibitively high GD spreads [21]. With any fiber design, in the absence of mode coupling, the GD spread accumulates linearly with fiber length, while strong mode coupling causes the GD spread to accumulate with the square-root of fiber length [19]. Nevertheless, even assuming low uncoupled GD spread and strong mode coupling, the GD spread caused by MD in MMF is far larger than that caused by PMD in SMF [23].

Most MDM demonstrations to date have used signal processing approaches generalized from SMF systems. Experiments [7], [24], [25] with two mode groups have employed 6×6 MIMO TDE adapted using data-aided methods, such as LMS [26], and implemented in non-real-time computation. In MDM, however, the computational complexity of MIMO TDE is known to be prohibitively high [19], [27], [28]. Computational

Manuscript received July 21, 2013; revised November 14, 2013 and January 15, 2014; accepted January 23, 2014. Date of publication January 27, 2014; date of current version April 27, 2014. This work was supported by National Science Foundation under Grant ECCS-1101905, Corning, Inc., and a Stanford Graduate Fellowship.

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

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

Digital Object Identifier 10.1109/JLT.2014.2303079

complexity per symbol scales linearly in D and in the GD spread from MD, while total hardware complexity scales quadratically in those parameters [19]. Moreover, blind equalization methods commonly used in SMF systems can perform poorly in MDM systems with MD and mode coupling [29].

Because of large values of D and large GD spreads, MDM systems ultimately must use different signal processing approaches than SMF systems to ensure tolerable complexity. As shown in [19], [27], and [30], MIMO FDE can reduce complexity, enabling the computational complexity per symbol to scale sublinearly in D and in the GD spread from MD (assuming a known channel), while hardware complexity still scales superlinearly in those parameters [19]. In long-haul MDM systems, mechanical or acoustic perturbations of installed MMFs are expected to cause mode coupling variations on time scales as short as tens of μs [31], similar to polarization coupling variations in SMF [12], [13]. These temporal variations demand adaptive MIMO FDEs that achieve fast adaptation, low output symbol-error ratio (SER), high throughput efficiency, and tolerable computational complexity.

Adaptive FDE techniques have been studied for wideband single-carrier MIMO wireless communications [32]–[35]. While those techniques may be relevant for MDM systems, MDM systems exhibit several important differences affecting the performance and feasibility of candidate techniques [36]. MDM systems have far higher GD spreads than typical wireless MIMO systems, when measured in units of symbol intervals. MDM MIMO channels may be near-unitary if mode-dependent loss and gain (MDL) is well-controlled [37], in contrast to wireless systems, where multipath fading typically causes MIMO channels to be far from unitary. MDM channels have symbol rates of tens of Gbaud, favoring simple techniques such as linear equalization, in contrast to wireless MIMO, where more complex nonlinear equalization techniques can be considered [38], [39].

In this paper, we study adaptive MIMO FDE for long-haul MDM systems using two candidate algorithms: LMS and recursive least squares (RLS). We study the tradeoffs between computational complexity, throughput efficiency, adaptation speed and output SER, and the impact of the fast Fourier transform (FFT) block length on them. For D ranging from 6 to 30, while either RLS or LMS can adapt to an unknown fiber within microseconds; RLS achieves faster adaptation, higher throughput efficiency, lower output SER, and greater tolerance to MDL, at the cost of moderately higher computational complexity per FFT block. Apart from adaptive algorithm design, low complexity and fast adaptation require the MDM system to have low GD spread. To this end, we use a family of graded-index MMFs in which uncoupled GD spread decreases with an increasing number of modes. Some results from this paper are summarized in a tutorial [36] on MDM systems, channel models and signal processing architectures. This paper presents a much more detailed performance and complexity analysis than [36].

The remainder of this paper is as follows. Section II presents a linear MIMO channel model for MDM systems and describes MIMO FDE. Section III reviews the LMS and RLS algorithms for FDE and their important properties. Section IV describes an exemplary long-haul MDM system, including a new MMF design optimized for low GD spread, and presents numerical simulations evaluating the complexity and performance of adaptive MIMO FDE methods. Sections V and VI provide discussion and conclusions, respectively.

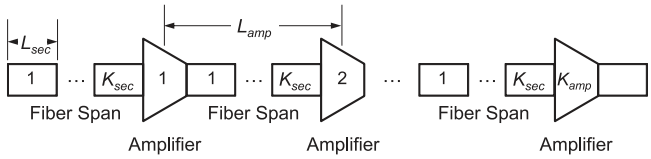

Fig. 1. Multi-section model of long-haul MDM system.

II. MDM System Model

A. Multi-Section Propagation Model

In order to study the effect of mode coupling on system performance and complexity, we model a long-haul MDM system as the concatenation of numerous short sections, each slightly longer than the length over which complex baseband modal fields remain correlated [28]. By decreasing the section length, thus increasing the number of sections, we can increase the strength of mode coupling. Throughout this paper, we assume the strong-coupling regime, so the number of independent sections is large compared to unity.

As shown in Fig. 1, the system is composed of K_{amp} spans, each comprising a fiber of length L_{amp}, followed by an amplifier to compensate for the mode-averaged loss of the fiber. Each span is subdivided into K_{sec} sections, each of length L_{sec}. The overall system has $K_{\mathrm{tot}} = K_{\mathrm{amp}} K_{\mathrm{sec}}$ sections and total length $L_{\mathrm{tot}} = K_{\mathrm{amp}} L_{\mathrm{amp}} = K_{\mathrm{amp}} K_{\mathrm{sec}} L_{\mathrm{sec}}$.

Assuming a MMF supporting D orthogonal propagating modes, excluding nonlinearity and noise, we represent the propagation operator as a $D \times D$ matrix multiplying complex baseband modal envelopes at frequency Ω [28]:

$$\mathbf{M}_{\mathrm{tot}}(\Omega) = \exp\left(-\frac{j}{2}\Omega^2 \bar{\beta}_2 L_{\mathrm{tot}}\right) \cdot \mathbf{M}_{\mathrm{MD}}(\Omega). \tag{1}$$

The exponential factor represents mode-averaged propagation, where $\bar{\beta}_2$ is the mode-averaged CD per unit length (for simplicity, we neglect the mode-averaged GD). The $D \times D$ matrix $\mathbf{M}_{\mathrm{MD}}(\Omega)$ represents mode-dependent effects, including MDL, MD and mode coupling (for simplicity, we neglect mode-dependent CD). It is written as a product of factors for each of the K_{amp} spans:

$$\mathbf{M}_{\mathrm{MD}}(\Omega) = \prod_{k=1}^{K_{\mathrm{amp}}} \mathbf{M}_{\mathrm{MD}}^{(k)}(\Omega), \tag{2}$$

where $\mathbf{M}_{\mathrm{MD}}^{(k)}(\Omega)$ represents mode-dependent effects in the kth span. This is given by

$$\mathbf{M}_{\mathrm{MD}}^{(k)}(\Omega) = \mathrm{diag}\left[\exp\left(\frac{g_1^{(k)}}{2}\right) \quad \cdots \quad \exp\left(\frac{g_D^{(k)}}{2}\right)\right] \cdot \prod_{l=1}^{K_{\mathrm{sec}}} \mathbf{V}^{(k,l)} \mathbf{\Lambda}(\Omega) \mathbf{U}^{(k,l)^H}, \tag{3}$$

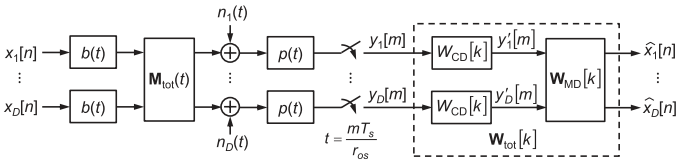

Fig. 2. Complex baseband model of MDM transmission with MIMO equalizer.

where ^{H} denotes Hermitian conjugate. The $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are frequency-independent random unitary matrices representing mode coupling in the lth section of the kth span.

We assume MDL from the transmission fibers is negligible compared to that from the optical amplifiers [40]–[42]. Each amplifier has uncoupled modal gains $g_i^{(k)}, i = 1, \ldots, D$, measured in dB or log-power-gain units. They satisfy $g_1^{(k)} + \cdots + g_D^{(k)} = 0$ and have root-mean-square (rms) value σ_g. In the strong-coupling regime with $K_{\text{amp}} \gg 1$ independent MDL sources, the statistics of coupled MDL are determined by the rms accumulated MDL $\xi = \sqrt{K_{\text{amp}}}\sigma_g$ [43].

In (3), $\mathbf{\Lambda}(\Omega)$ represents the uncoupled MD in one section, which is assumed to be the same in all sections. It is given by

$$\mathbf{\Lambda}(\Omega) = \text{diag}\left[\exp\left(-j\Omega\tau_1\right) \quad \ldots \quad \exp\left(-j\Omega\tau_D\right)\right], \quad (4)$$

where $\tau_i, i = 1, \ldots, D$ are the uncoupled modal GDs, which satisfy $\tau_1 + \cdots + \tau_D = 0$. They have rms value $\sigma_\tau = \Delta\beta_{1,\text{rms}} L_{\text{sec}}$, where $\Delta\beta_{1,\text{rms}}$ is the rms uncoupled MD per unit length of the MMF. In the strong-coupling regime with $K_{\text{tot}} >> 1$ independent sections, the statistics of the coupled GDs are determined by the rms coupled GD $\sigma_{gd} = \sqrt{K_{\text{tot}}}\sigma_\tau = \Delta\beta_{1,\text{rms}}\sqrt{L_{\text{sec}} L_{\text{tot}}}$ [44].

B. Discrete-Time System Model

Neglecting nonlinearity and assuming digital coherent detection with perfect carrier recovery, an MDM transmission system can be described by the complex baseband system shown in Fig. 2. The input in the jth mode in the nth symbol interval, $x_j[n]$, is a complex-valued symbol from a particular constellation. Transmitter pulse shaping and electrical-to-optical conversion is represented by $b(t)$. Linear propagation is described by a $D \times D$ channel impulse response matrix $\mathbf{M}_{\text{tot}}(t)$, which is the inverse Fourier transform (FT) of (1), in which $m_{dj}(t)$ is the impulse response between the dth input mode and the jth output mode. The dominant noise source is amplified spontaneous emission from inline amplifiers [1]. The additive noises $n_j(t), j = 1, \ldots, D$ are modeled as spectrally white with power spectral density $N_0/2$ over the signal bandwidth [1], and can be modeled accurately as spatially white, provided K_{amp} is sufficiently large [43]. After the addition of noise, optical-to-electrical conversion and electrical filtering are performed, represented by $p(t)$. We define an overall impulse response between the dth input mode and the jth output mode

$$q_{dj}(t) = b(t) * m_{dj}(t) * p(t) \quad (5)$$

and a filtered noise $n'_j(t) = p(t) * n_j(t)$.

The output of each filter is sampled at rate r_{os}/T_s, where T_s is the symbol duration and r_{os} is the receiver oversampling ratio [15]. The output in the jth mode at the mth sampling instant $t = mT_s/r_{os}$ is given by

$$y_j[m] = \sum_n \sum_{d=1}^{D} x_j[n]\, q_{dj}\left(\frac{mT_s}{r_{os}} - nT_s\right) + n'_j\left(\frac{mT_s}{r_{os}}\right). \quad (6)$$

The discrete-time Fourier transforms (DTFTs) of the output sample sequences $y_j[m], j = 1, \ldots, D$, can be expressed as a $D \times 1$ vector

$$\begin{bmatrix} \sum_m y_1[m]\, e^{-j\omega m} \\ \vdots \\ \sum_m y_D[m]\, e^{-j\omega m} \end{bmatrix} = \begin{bmatrix} \sum_l Y_1\left(\frac{(\omega - 2\pi l)\, r_{os}}{T_s}\right) \\ \vdots \\ \sum_l Y_D\left(\frac{(\omega - 2\pi l)\, r_{os}}{T_s}\right) \end{bmatrix} \approx \begin{bmatrix} Y_1\left(\frac{\omega r_{os}}{T_s}\right) \\ \vdots \\ Y_D\left(\frac{\omega r_{os}}{T_s}\right) \end{bmatrix}. \quad (7)$$

The components in the rightmost term of (7) are given by

$$\begin{bmatrix} Y_1(\Omega) \\ \vdots \\ Y_D(\Omega) \end{bmatrix} = B(\Omega)\mathbf{M}_{\text{tot}}(\Omega)P(\Omega) \begin{bmatrix} X_1\left(e^{jT_s\Omega}\right) \\ \vdots \\ X_D\left(e^{jT_s\Omega}\right) \end{bmatrix} + \begin{bmatrix} N'_1(\Omega) \\ \vdots \\ N'_D(\Omega) \end{bmatrix}, \quad (8)$$

where $B(\Omega), \mathbf{M}_{\text{tot}}(\Omega), P(\Omega), N'_j(\Omega)$ are the FTs of $b(t), \mathbf{M}_{\text{tot}}(t), p(t), n'_j(t)$, respectively, and $X_j\left(e^{j\omega}\right)$ is the DTFT of $x_j[n]$. In computing the FT of $n'_j(t)$ and the DTFT of $x_j[n]$, we consider any particular finite-length realizations of these random processes to be finite-energy signals. The approximation in (7) assumes $P(\Omega)$ blocks all components above the Nyquist frequency $r_{os}/2T_s$, preventing aliasing.

C. Frequency-Domain Equalization

The optimization and complexity of non-adaptive FDEs for MDM, assuming the channel is known *a priori*, were addressed in [19]. An equalizer for CD [14], [19] must accommodate a delay spread of $N_{\text{CD}} = \left\lceil 2\pi \left|\bar{\beta}_2\right| L_{\text{tot}} \left(r_{os} R_s\right)^2 \right\rceil$ samples, where $R_s = 1/T_s$ is the symbol rate and $\lceil x \rceil$ denotes the ceiling function. Assuming strong mode coupling, an equalizer for MD [19] must accommodate a delay spread of $N_{\text{MD}} = \left\lceil \sqrt{K_{\text{tot}}}\sigma_\tau u_D(p)\, r_{os} R_s \right\rceil = \left\lceil \Delta\beta_{1,\text{rms}}\sqrt{L_{\text{sec}} L_{\text{tot}}}\, u_D(p)\, r_{os} R_s \right\rceil$ samples, where $u_D(p)$ is defined such that $\sigma_{gd} u_D(p)$ is no shorter than the coupled GD spread with probability $1 - p$. For typical values of D and $p \sim 10^{-4}$ to 10^{-6}, $u_D(p) \sim 4$ to 5 [19], [45]. A combined equalizer for CD and MD must accommodate a delay spread of $N_{\text{CD}} + N_{\text{MD}}$ samples.

The MDM channel corresponds to linear convolution of arbitrary-length sequences in the time domain (6) or multiplication of DTFTs in the continuous frequency domain (7) and (8). Efficient realization of an FDE relies on using the discrete Fourier transform (DFT), implemented by an FFT of block length N_{FFT}, for conversion between time and frequency, while being able to represent the MDM channel as circular convolution of finite-length sequences in the time domain, corresponding to multiplication in the discrete frequency domain.

There are two well-known approaches for enabling FFT-based FDEs.

One approach is to use block convolution, e.g., overlap-save convolution, as in [46], [47]. This avoids any overhead associated with a cyclic prefix, but complicates realization of an adaptive FDE. A constrained adaptive FDE requires additional FFTs to enforce time-domain gradient constraints [46]. An unconstrained adaptive FDE avoids these additional FFTs, but exhibits slower adaptation and higher excess error [46].

A second approach, adopted here, is to prepend a cyclic prefix of length N_{CP} to each block of N_{FFT}/r_{os} symbols before transmission, as in systems using orthogonal frequency-division multiplexing [48]. At the receiver, the first N_{CP} samples of each received block of $N_{\mathrm{FFT}}+N_{\mathrm{CP}}$ samples are discarded, and the remaining N_{FFT} samples can be processed independent of other blocks. The cyclic prefix length N_{CP} must be no shorter than the channel delay spread ($N_{\mathrm{CD}}, N_{\mathrm{MD}}$ or $N_{\mathrm{CD}}+N_{\mathrm{MD}}$). A drawback of this approach is a reduction of throughput and average-power efficiency, which is quantified by a cyclic prefix efficiency parameter

$$\eta_{\mathrm{CP}} = \frac{N_{\mathrm{FFT}}}{N_{\mathrm{FFT}}+N_{\mathrm{CP}}}. \tag{9}$$

Given a channel delay spread defining N_{CP}, efficiency is maximized by choosing $N_{\mathrm{FFT}} >> N_{\mathrm{CP}}$. Since the cyclic prefix simplifies realization of an adaptive FDE [46], this approach has become popular in wireless systems [32]–[34].

The MDM channel frequency-domain relationship when using a cyclic prefix is obtained by sampling (8) at N_{FFT} equally spaced frequencies $\Omega = 2\pi r_{os}\left(k-N_{\mathrm{FFT}}/2\right)/N_{\mathrm{FFT}}T_s, k = 0, ..., N_{\mathrm{FFT}}-1$

$$\begin{aligned}\begin{bmatrix} Y_1[k] \\ \vdots \\ Y_D[k] \end{bmatrix} &= B\left(\frac{2\pi r_{os}\left(k-N_{\mathrm{FFT}}/2\right)}{N_{\mathrm{FFT}}T_s}\right) \\ &\cdot \mathbf{M}_{\mathrm{tot}}\left(\frac{2\pi r_{os}\left(k-N_{\mathrm{FFT}}/2\right)}{N_{\mathrm{FFT}}T_s}\right) \\ &\cdot P\left(\frac{2\pi r_{os}\left(k-N_{\mathrm{FFT}}/2\right)}{N_{\mathrm{FFT}}T_s}\right) \\ &\cdot \begin{bmatrix} X_1[k] \\ \vdots \\ X_D[k] \end{bmatrix} + \begin{bmatrix} N_1'[k] \\ \vdots \\ N_D'[k] \end{bmatrix} \\ &= \mathbf{Q}[k]\cdot \begin{bmatrix} X_1[k] \\ \vdots \\ X_D[k] \end{bmatrix} + \begin{bmatrix} N_1'[k] \\ \vdots \\ N_D'[k] \end{bmatrix}. \end{aligned} \tag{10}$$

The components of $(Y_1[k],\ldots,Y_D[k])^T$, $(X_1[k], \ldots, X_D[k])^T$ and $(N_1'[k],\ldots,N_D'[k])^T$ are N_{FFT}-point DFTs of blocks of the time-domain signals in (6).

Considering a non-adaptive FDE [19], computational complexity is minimized by using a single $D\times D$ matrix FDE $\mathbf{W}_{\mathrm{tot}}[k]$ to compensate for the CD, MD and other mode-dependent effects described by $\mathbf{M}_{\mathrm{tot}}(\Omega)$, which requires a prefix length $N_{\mathrm{CD}}+N_{\mathrm{MD}}$. However, to facilitate realization of a fast-adapting FDE for MD, we compensate CD by a bank of D static scalar equalizers, each denoted by $W_{\mathrm{CD}}[k]$. The notation implies they are FDEs, but they could be realized using any of the methods in [17], [49]–[51]. They are followed by an adaptive $D\times D$ matrix FDE $\mathbf{W}_{\mathrm{MD}}[k]$ to compensate for the MD (and other mode-dependent effects) described by $\mathbf{M}_{\mathrm{MD}}(\Omega)$, which requires cyclic prefix length N_{MD}. As shown in Fig. 2, the cascade of the CD and MD equalizers is equivalent to a combined equalizer $\mathbf{W}_{\mathrm{tot}}[k] = W_{\mathrm{CD}}[k]\,\mathbf{W}_{\mathrm{MD}}$. The equalizer output is given by

$$\begin{bmatrix} \hat{X}_1[k] \\ \vdots \\ \hat{X}_D[k] \end{bmatrix} = \mathbf{W}_{\mathrm{tot}}[k]\begin{bmatrix} Y_1[k] \\ \vdots \\ Y_D[k] \end{bmatrix} = \mathbf{W}_{\mathrm{MD}}[k]\begin{bmatrix} Y_1'[k] \\ \vdots \\ Y_D'[k] \end{bmatrix}, \tag{11}$$

where $[Y_1'[k],\ldots,Y_D'[k]]^T$ represents the outputs of the CD equalizers.

III. Adaptive Frequency-Domain Equalization

At each discrete frequency $k = 0, ..., N_{\mathrm{FFT}}-1$, the error signal is given by

$$\begin{aligned}\mathbf{e}[k] &= \tilde{\mathbf{x}}[k] - \mathbf{W}_{\mathrm{MD}}[k]\,\tilde{\mathbf{y}}[k] \\ &= \tilde{\mathbf{x}}[k] - \mathbf{W}_{\mathrm{MD}}[k]\left(\mathbf{Q}[k]\,\tilde{\mathbf{x}}[k] + \tilde{\mathbf{n}}'[k]\right)\end{aligned} \tag{12}$$

where $\tilde{\mathbf{x}}[k] = [X_1[k]\ldots X_D[k]]^T$ represents a block of known or estimated data symbols, $\tilde{\mathbf{y}}[k] = [Y_1'[k]\ldots Y_D'[k]]^T$ represents a block of samples at the CD equalizer outputs, $\tilde{\mathbf{n}}'[k] = [N_1'[k]\ldots N_D'[k]]^T$ represents the filtered noise and $\mathbf{Q}[k]$ is the overall channel transfer function matrix from (10). Minimization of the total frequency-domain signal error $\sum_{k=0}^{N_{\mathrm{FFT}}-1} E\{\mathbf{e}[k]^H\,\mathbf{e}[k]\}$ is equivalent to separately minimizing the error terms at each frequency, as the equalizer matrices $\mathbf{W}_{\mathrm{MD}}[k]$ can be optimized independently at each k.

For a known channel, $E\left\{\mathbf{e}[k]^H\,\mathbf{e}[k]\right\}$ is minimized by the linear minimum mean square error (MMSE) filter [52]:

$$\mathbf{W}_{\mathrm{MD}}[k] = \left(\mathbf{Q}[k]^H R_{\tilde{\mathbf{n}}'}[k]^{-1}\mathbf{Q}[k] + \frac{\mathbf{I}}{P_x}\right)^{-1}\mathbf{Q}[k]^H\,R_{\tilde{\mathbf{n}}'}[k]^{-1} \tag{13}$$

where $P_x = E\{|x_j[n]|^2\}$ is the average transmitted power in each mode, $\mathbf{I}$ is a $D\times D$ identity matrix and $R_{\tilde{\mathbf{n}}'}[k] = E\{\tilde{\mathbf{n}}'[k]\,\tilde{\mathbf{n}}'[k]^H\}$ is the autocorrelation matrix of filtered noise at discrete frequency k.

For an unknown channel, since $\mathbf{Q}[k]$ are not known *a priori* at the equalizer, the $\mathbf{W}_{\mathrm{MD}}[k]$ must be computed iteratively. Here, we consider equalizer adaptation using the LMS and RLS algorithms.

LMS is a stochastic gradient descent minimization using instantaneous estimates of the error $\mathbf{e}[k]$ [46], [53] described by an update equation

$$\mathbf{W}_{\mathrm{MD}}[k] \leftarrow \mathbf{W}_{\mathrm{MD}}[k] + \left(\tilde{\mathbf{x}}[k] - \mathbf{W}_{\mathrm{MD}}[k]\,\tilde{\mathbf{y}}[k]\right)\tilde{\mathbf{y}}[k]^H\,\mu. \tag{14}$$

The convergence rate and performance of LMS depends on the scalar step size μ, which for convergence must satisfy

TABLE I
COMPUTATIONAL COMPLEXITY PER BLOCK OF LENGTH N_{FFT}, WHICH CONVEYS $D \cdot N_{\mathrm{FFT}}/r_{\mathrm{os}}$ SYMBOLS

Operation	Number of complex multiplications	Number of complex additions
DFT or inverse DFT	$D{\cdot}N_{\mathrm{FFT}}\log_2(N_{\mathrm{FFT}})/2$	$D{\cdot}N_{\mathrm{FFT}}\log_2(N_{\mathrm{FFT}})$
Equalization (11)	$N_{\mathrm{FFT}}{\cdot}D^2$	$N_{\mathrm{FFT}}{\cdot}(D^2{-}D)$
Adaptation with LMS (14)	$N_{\mathrm{FFT}}{\cdot}(D^2{+}D)$	$N_{\mathrm{FFT}}{\cdot}(D^2{+}D)$
Adaptation with RLS (15) and (16)	$N_{\mathrm{FFT}}{\cdot}(5D^2{+}2D)$	$N_{\mathrm{FFT}}{\cdot}(4D^2)$

$0 < \mu < 2/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of the autocorrelation matrix of $\tilde{\mathbf{y}}[k]$, $R_{\tilde{\mathbf{y}}}[k] = E\left\{\tilde{\mathbf{y}}[k]\,\tilde{\mathbf{y}}[k]^H\right\}$ [46].

RLS involves iterative minimization of an exponentially weighted cost function, treating the minimization problem as deterministic [46], [53]. It is described by update equations

$$\mathbf{W}_{\mathrm{MD}}[k] \leftarrow \mathbf{W}_{\mathrm{MD}}[k] + \left(\tilde{\mathbf{x}}[k] - \mathbf{W}_{\mathrm{MD}}[k]\,\tilde{\mathbf{y}}[k]\right)\tilde{\mathbf{y}}[k]^H\left(\mathbf{R}[k]\,\kappa^{-1}\right) \tag{15}$$

$$\mathbf{R}[k] \leftarrow \left(\mathbf{R}[k]\,\kappa^{-1}\right) - \frac{\left(\mathbf{R}[k]\,\kappa^{-1}\right)\tilde{\mathbf{y}}[k]\,\tilde{\mathbf{y}}[k]^H\left(\mathbf{R}[k]\,\kappa^{-1}\right)}{1+\tilde{\mathbf{y}}[k]^H\left(\mathbf{R}[k]\,\kappa^{-1}\right)\tilde{\mathbf{y}}[k]}. \tag{16}$$

Here, $\mathbf{R}[k]$ is a $D \times D$ tracked inverse time-averaged weighted correlation matrix at frequency k [32], [34], which is initialized with the identity matrix times a large positive number, and κ is a forgetting factor satisfying $0 << \kappa < 1$. The version of the RLS algorithm given above uses the matrix inversion lemma in (16) for lower-complexity computation of $\mathbf{R}[k]$, which becomes more important for higher values of D.

Assuming n_{tr} blocks of known or estimated symbols are required for training until convergence, and including the cyclic prefix, the total time required to adapt an FDE is

$$T_{\mathrm{adapt}} = \frac{n_{tr}\left(N_{\mathrm{FFT}} + N_{\mathrm{CP}}\right)T_s}{r_{os}}. \tag{17}$$

Typically T_s, r_{os} and $N_{\mathrm{CP}} = N_{\mathrm{MD}}$ are given system parameters. The parameter n_{tr} is determined mainly by the algorithm and convergence criteria. Hence, N_{FFT} is the major parameter that may be decreased to minimize the training time T_{adapt}. However, given a prefix length N_{CP}, decreasing N_{FFT} reduces the cyclic prefix efficiency (9). In Section IV, we will see that it can be necessary to reduce N_{MD} in order to minimize T_{adapt} while maintaining high efficiency.

As rough measures of implementation complexities, we evaluate the number of complex multiplications and additions required for the LMS and RLS algorithms. Table I gives the computational complexities of the algorithms for a training block of length N_{FFT}, considering the steps (13)–(16) (see Appendix A for the details of the derivations). To evaluate the total computational complexity to adapt an FDE, all the complexities in Table I should be multiplied by the number of training blocks n_{tr}.

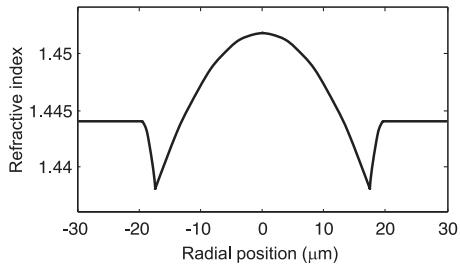

Fig. 3. Optimized fiber index profile for $D = 12$ modes.

The computational complexity per data-bearing symbol is obtained by dividing the complexities in Table I by the number of symbols per block, $D \cdot N_{\mathrm{FFT}}/r_{os}$. The per-symbol complexities scale at most linearly with D. They depend on N_{FFT} only through the DFT and inverse DFT operations, which favor the choice of small N_{FFT} to minimize complexity. For values of D large relative to $\log_2(N_{\mathrm{FFT}})$, the complexity of adaptation and equalization dominates over the complexity of DFT/inverse DFT operations, and the dependence of complexity on N_{FFT} is reduced.

IV. FDE COMPLEXITY AND PERFORMANCE EXAMPLES

In this section, we study the complexity, adaptation time, and SER performance of MIMO FDEs using LMS and RLS algorithms. The system end-to-end GD spread has a major effect on complexity and performance, so we propose a novel fiber design having very low uncoupled MD.

A. Multimode Fiber Design

Managing an MDM system's end-to-end GD spread is important in controlling the computational and hardware complexity of the MIMO FDE [19], [28] and, as shown here, in achieving fast adaptation of the MIMO FDE. Here, we propose to use fibers with low uncoupled GD spread and rely on strong mode coupling, induced by splices or other perturbations, to further reduce the GD spread [19], [20].

We consider a family of "graded-index graded depressed-cladding" (GIGDC) index profiles, inspired by [54]–[56], in which a parabolic core index profile is extended smoothly into a depressed cladding. Fig. 3 shows a profile optimized for $D = 12$ modes, and Table II gives parameters for fibers supporting $D = 6, 12, 20$ and 30 modes. These have been computed by numerical solution of the vector wave equation, without assumption of weak guidance [21]. For each value of D, the numerical aperture is $NA = 0.150$, and the core radius is chosen so the number of propagating modes is exactly D over the C band, and increases at wavelengths just below 1530 nm. This approach optimizes confinement of propagating modes, minimizing bending losses, mode-dependent losses and mode-dependent CD. As included in Table II, the mode-averaged effective areas scale as $D^{0.78}$ and minimum modal effective areas scale as $D^{0.43}$. Most relevant here, the rms MD $\Delta\beta_{1,\mathrm{rms}}$ tends to decrease as D increases from 6 to 30. This behavior, while perhaps counterintuitive, may be justified by observing that as the core radius is increased

TABLE II
UNCOUPLED PARAMETERS OF GRADED-INDEX GRADED-DEPRESSED-CLADDING FIBERS FOR DIFFERENT NUMBERS OF MODES, ASSUMING $NA = 0.150$ AND $\lambda = 1550$ nm

Number of modes D	6	12	20	30
Core radius (μm)	9.8	13.1	16.4	19.7
Mode-averaged effective area $\overline{A}_{eff}$ (μm^2)	170	265	427	590
Minimum modal effective area $A_{eff,\,min}$ (μm^2)	102	136	170	204
Mode-averaged chromatic dispersion $\overline{\beta}_2$ (ps^2/km)	−26.2	−26.7	−26.8	−26.9
Root-mean-square modal dispersion $\Delta\beta_{1,rms}$ (ps/km)	34	29	9	10

to support more modes, the index profile "seen" by the modes increasingly resembles an infinite parabola, which is ideally free of MD to first-order [22]. For GIGDC fibers, the mode-dependent CD has an rms value of only 3% of the mode-averaged CD $\bar{\beta}_2$, justifying our neglect of mode-dependent CD in (1). Low mode-dependent CD ensures that a low GD spread can be achieved over a wide range of wavelengths.

B. Transmission System

We consider a long-haul fiber system described by the multi-section model of Section II-A (see Fig. 1). The system has $K_{\mathrm{amp}} = 20$ spans, each of length $L_{\mathrm{amp}} = 100$ km, for a total length $L_{\mathrm{tot}} = 2000$ km. The family of GIGDC fibers with uncoupled parameters given in Table II is assumed. To vary the strength of mode coupling, the number of sections per span K_{sec} is varied from 1 to 100, corresponding to a section length L_{sec} from 100 to 1 km, and an rms end-to-end GD spread $\sigma_{gd} = \Delta\beta_{1,\mathrm{rms}}\sqrt{L_{\mathrm{sec}}L_{\mathrm{tot}}}$. When MDL is present, the uncoupled modal gains in each span are uniformly distributed with rms value σ_g, resulting in an rms accumulated MDL $\xi = \sqrt{K_{\mathrm{amp}}}\sigma_g$, measured in decibal.

MDM transmission is described by the system model of Section II-B. A symbol rate $R_s = 1/T_s = 32$ Gbaud and a receiver oversampling rate $r_{os} = 2$ are assumed. The transmitter pulse shape $b(t)$ is a rectangular pulse of duration T_s filtered by a fifth-order Bessel lowpass filter with -3-dB cutoff frequency $0.8/T_s$, while the receiver filter $p(t)$ is a fifth-order Butterworth lowpass filter with -3-dB cutoff frequency $0.4r_{os}/T_s$ [15]. Transmitted symbols are drawn from a quadrature phase-shift keying constellation with average power P_x. The SNR as the transmitted signal power over received noise per mode is defined as $\mathrm{SNR} = P_x/\sigma_N^2$, where $\sigma_N^2 = N_0 R_s r_{os}$. We initially neglect phase noise and discuss its effects in Section V.

C. Computational Complexity and Cyclic Prefix Efficiency

We first study the computational complexity and cyclic prefix efficiency of the adaptive MIMO FDE. These depend strongly on the MD delay spread N_{MD}, which we compute as described in Section II-C, assuming $p = 10^{-5}$ and neglecting MDL ($\xi =$ 0 dB). The presence of significant MDL might change N_{MD} slightly, as studied for $D = 2$ [57].

TABLE III
IMPULSE RESPONSE DURATIONS (IN SAMPLES) FOR DIFFERENT SECTION LENGTHS, ASSUMING $L_{tot} = 2000$ km, $R_s = 32$ Gbaud, $r_{os} = 2$, $p = 10^{-5}$ AND $\xi = 0$ dB

Number of modes D		6	12	20	30
Chromatic dispersion N_{CD}		1,349	1,375	1,377	1,381
Modal dispersion N_{MD}	L_{sec} = 100 km	4,774	3,852	1,109	1,312
	L_{sec} = 10 km	1,510	1,218	351	415
	L_{sec} = 1 km	477	385	111	131

TABLE IV
MIMO EQUALIZER: CYCLIC PREFIX LENGTHS, BLOCK LENGTHS, CYCLIC PREFIX EFFICIENCIES AND COMPUTATIONAL COMPLEXITIES, ASSUMING THE SAME PARAMETERS AS TABLE III AND $L_{\mathrm{sec}} = 1$ km

Number of modes D		6	12	20	30
Cyclic prefix length $N_{CP} = N_{MD}$		477	385	111	131
Block length N_{FFT}		2^{11}	2^{11}	2^9	2^9
Cyclic prefix efficiency η_{CP} (%)		81.1	84.1	82.1	79.6
Known channel	Complex multiplications/symbol	34	46	58	78
	Complex additions/symbol	54	66	74	94
LMS-adapted, in training block	Complex multiplications/symbol	48	72	100	140
	Complex additions/symbol	68	92	116	156
RLS-adapted, in training block	Complex multiplications/symbol	98	170	262	382
	Complex additions/symbol	102	162	234	334

Table III gives values of N_{MD} for $D = 6$, 12, 20 and 30 modes for section lengths $L_{\mathrm{sec}} = 100$, 10 and 1 km, comparing these to the CD delay spread N_{CD}. In order for the computational complexity of an adaptive FDE for MD to be roughly comparable to that for CD, a system should be designed so that N_{MD} is several times smaller than N_{CD}. This requirement can be traced to the adaptivity and increased dimensionality of the $D \times D$ matrix FDE for compensating MD, as compared to the D static scalar FDEs for compensating CD. Although an arbitrarily small section length is possible in principle, we believe a realistic target is $L_{\mathrm{sec}} = 1$ km, and this value is assumed hereafter. For this choice, N_{MD} decreases from 477 to 131 (2.8 to 10.5 times smaller than respective values of N_{CD}) as D ranges from 6 to 30.

Given these values of N_{MD}, MIMO FDEs are designed using the parameters given in Table IV. Prefix lengths are chosen to satisfy $N_{\mathrm{CP}} = N_{\mathrm{MD}}$. Block lengths N_{FFT} varying from 2^{11} to 2^9 as D ranges from 6 to 30 are chosen to ensure fast adaptation (see below). Corresponding values of the prefix efficiency η_{CP}, given by (9), are roughly 80% for all values of D. As explained later, RLS adaptation may be sufficiently fast to permit an increase in N_{FFT}, which increases η_{CP}.

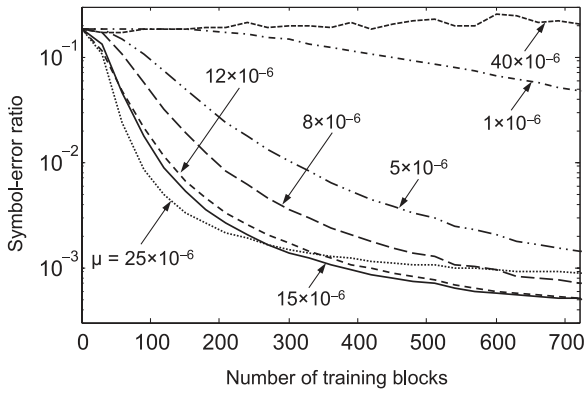

Fig. 4. Average symbol-error ratio versus number of training blocks for different values of μ for LMS-adapted FDE, assuming $D = 12$ modes, SNR = 10.5 dB, $\xi = 0$ dB and $D = 12$ modes.

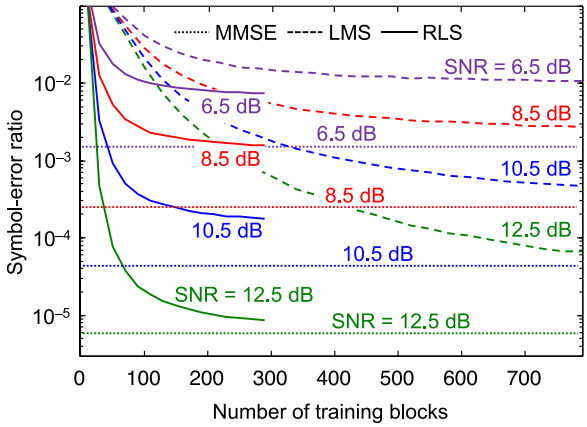

Fig. 5. Average symbol-error ratio versus number of training blocks for SNR = 6.5, 8.5, 10.5 and 12.5 dB with MMSE filtering (dotted lines), RLS (solid lines) and LMS (dashed lines) algorithms assuming $\kappa = 0.999$, $\mu = 1.5 \times 10^{-5}$, $\xi = 0$ dB and $D = 12$ modes.

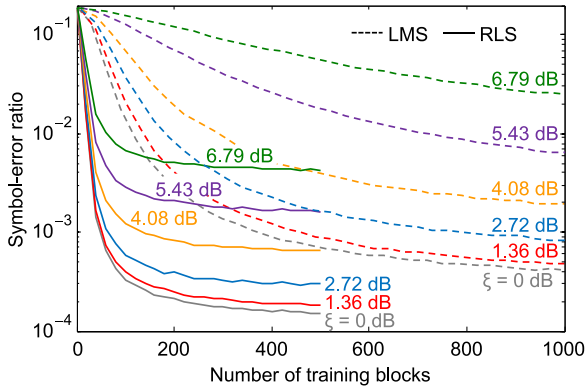

Fig. 6. Average symbol-error ratio versus number of training blocks for accumulated MDL $\xi = 0$, 1.36, 2.72, 4.08, 5.43 and 6.79 dB with RLS algorithm (solid lines) with $\kappa = 0.999$ and LMS algorithm (dashed lines) with $\mu = 1.5 \times 10^{-5}$, 1.3×10^{-5}, 1.1×10^{-5}, 9.3×10^{-6}, 5.1×10^{-6} and 3.1×10^{-6} respectively, assuming SNR = 10.5 dB and $D = 12$ modes.

Table IV gives values of computational complexity per data symbol for MIMO FDE. "Known channel" is the complexity of equalization only, corresponding to the twice the first line plus the second line of Table I. "LMS-adapted" or "RLS-adapted" is the complexity of both equalization and adaptation, and also includes the third or fourth line of Table I. In Table IV, each of the above has been divided by $D \cdot N_{\mathrm{FFT}}/r_{os}$, the number of symbols per block.

As compared to equalizing a known channel, LMS adaptation increases the required complex multiplications per symbol by 1.4 to 1.8 and the required complex additions per symbol by 1.3 to 1.7 as D ranges from 6 to 30. RLS adaptation increases the required complex multiplications per symbol by 2.9 to 4.9 and the required complex additions per symbol by 1.9 to 3.6 over the same range of D. The complexities of LMS and RLS adaptation are compared further in Section V.

For comparison, static overlap-save based FDE of CD [19] requires about 33 complex multiplications per symbol and about 61 complex additions per symbol for all values of D, assuming a block length $N_{\mathrm{FFT}} = 2^{14}$.

D. Adaptation Time and SER Performance

We now study the adaptation time and SER performance of the MIMO FDE and their dependence on SNR, MDL and the number of modes. We present values of the SER as a function of the number of training blocks n_{tr} obtained by averaging Monte-Carlo simulations over random channel and symbol sequence realizations. We continue to assume a section length $L_{\mathrm{sec}} =$ 1 km. For the RLS algorithm, we choose a forgetting factor $\kappa = 0.999$, whose optimization has negligible effect on both the converged SER value and the adaptation time. For the LMS algorithm, the choice of the step size μ is critical. As shown in Fig. 4, small values of μ result in impractically slow adaptation times whereas large values of μ yield higher asymptotic SERs and can cause the LMS algorithm to violate the convergence condition discussed in Section III. Considering this trade-off, for MDL-free channels, we choose an optimal value of $\mu = 1.5 \times 10^{-5}$. In the presence of MDL, since the condition number of the MDM channel matrix $\mathbf{M}_{\mathrm{MD}}(\Omega)$ is increased [43], μ is reduced to optimize the convergence rate.

First we fix the number of modes to $D = 12$ and study adaptation in different regimes of SNR and MDL.

Fig. 5 shows SER versus n_{tr} achieved by RLS or LMS at different values of SNR per mode, assuming no MDL ($\xi = 0$ dB). As expected, RLS converges faster, and to a lower asymptotic SER, than LMS. The logarithm of the asymptotic SER scales roughly as $\mathrm{SNR}^{-1.5}$ for RLS and as $\mathrm{SNR}^{-1.3}$ for LMS, leading to an increasing disparity in their asymptotic SERs with increasing SNR. As the SNR increases, the time required for convergence to the asymptotic SER increases slowly for RLS, and more rapidly for LMS. As a comparison case, performance bound of the MMSE filter given by (13) is also shown in Fig. 5. The asymptotic SER of RLS approaches that of the MMSE equalizer much faster than that of LMS as the SNR increases. Over the range of SNRs considered, the SERs for using RLS and LMS are, respectively, ~2–6 times and ~6–10 times higher than those for the MMSE equalizer.

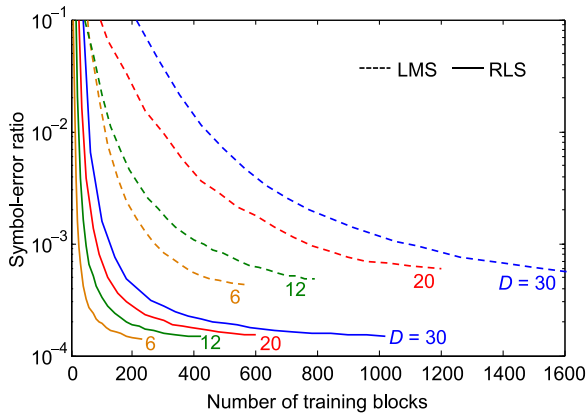

Fig. 7. Average symbol-error ratio versus number of training blocks for $D = 6$, 12, 20 and 30 modes with RLS (solid line) and LMS (dashed line) algorithms assuming $\kappa = 0.999$, $\mu = 1.5 \times 10^{-5}$, $\xi = 0$ dB and SNR = 10.5 dB.

TABLE V
VALUES OF ADAPTATION TIME T_{adapt} TO REACH VARIOUS SYMBOL-ERROR RATIOS USING RLS AND LMS ALGORITHMS FOR $D = 6$, 12, 20 AND 30 MODES, USING THE RESULTS OF FIG. 6

Adaptive Algorithm	Number of Modes D	T_{adapt} (μs)		
		SER = 2×10^{-3}	SER = 7×10^{-4}	SER = 2×10^{-4}
LMS	6	7.9	12.4	-
	12	10.3	21.0	-
	20	5.3	9.2	-
	30	7.8	13.5	-
RLS	6	0.7	1.1	3.9
	12	1.4	2.1	6.9
	20	0.6	1.0	3.0
	30	0.9	1.5	4.7

Fig. 6 shows the SER versus n_{tr} achieved by RLS or LMS at different values of the accumulated MDL ξ, assuming an SNR of 10.5 dB. As expected, RLS is more robust than LMS, converging faster to a lower asymptotic SER. The convergence rate of RLS is not adversely affected for the values of MDL considered, and the logarithm of the asymptotic SER increases roughly proportional to ξ. On the other hand, increasing MDL causes the convergence of LMS to slow down and causes the logarithm of the asymptotic SER to increase roughly proportional to ξ, but at a higher rate than for RLS. Note that for LMS, the step size μ is optimized for each value of ξ. If μ is held constant at the value optimal for the MDL-free case, even low-to-intermediate values of ξ can prevent convergence and cause outage.

Now we study adaptation as a function of the number of modes D. Fig. 7 shows SER versus n_{tr} achieved by RLS or LMS for different values of D, assuming an SNR of 10.5 dB and no MDL ($\xi = 0$ dB). As D increases, although the block length N_{FFT} decreases slightly, the MIMO equalizer dimension scales as D^2, so the number of training blocks required for convergence increases. For RLS, the number of training blocks needed for convergence is roughly proportional to D (the curves nearly overlap when scaled by D^{-1}), and the knee and flat-SER regions of the adaptation curves occur at about $6D$ and $15D$, respectively. For LMS, the adaptation curves do not overlap with any power-of-D scaling. The knees occur at about $20D$ to $30D$, while the flat-SER regime starts at about $50D$ to $70D$. The asymptotic SERs for RLS are around 2×10^{-4}, while those for LMS are a factor of $\sim$3 higher.

To quantify equalizer adaptation times, we extract from Fig. 7 values of n_{tr} such that target SER values are achieved, and compute T_{adapt} using (17). Table V presents these adaptation times for RLS and LMS for various numbers of modes D. Adaptation times for RLS range from roughly 1 to 7 μs, depending on D and the target SER, while adaptation times for LMS are roughly ten times longer for the same target SER for a given D. These implications of results are discussed in Section V.

V. DISCUSSION

As noted above, adaptive MIMO FDEs for MDM have several important design objectives, including scalability to many modes, reliable convergence and low asymptotic SERs in the presence of noise and MDL, fast initial adaptation and subsequent tracking, high cyclic prefix efficiency, and low computational complexity. We have found that RLS outperforms LMS in most key respects.

As the number of modes D increases, RLS maintains faster adaptation and lower asymptotic SERs than LMS (see Fig. 7). In the presence of noise, RLS achieves lower asymptotic SERs than LMS (see Fig. 5). In the presence of MDL, RLS converges more reliably and achieves lower asymptotic SERs than LMS (see Fig. 6), which is critical, because experiments [58] have demonstrated that MDL can readily cause the channel matrix to become ill-conditioned.

Reliable tracking of dynamic channel behavior is a critical requirement in optical transport networks. It has been estimated that MIMO channels in long-haul MDM systems will change on a time scale of 25 μs [31], similar to long-haul polarization-division-multiplexed systems in SMF. There do not yet exist models for the dynamic evolution of MDM channels, which would enable study of the tracking behavior of adaptive MIMO FDEs. For sake of discussion, we conservatively assume that in adapting to an unknown channel, an FDE must nearly reach its asymptotic SER in no more than 25 μs, which should ensure it can track dynamics in a long-haul MDM system.

Using the parameter values in Table V, values of the adaptation time T_{adapt} for the LMS algorithm are close to 25 μs, i.e., LMS appears fast enough to track long-haul MDM channels. Assuming these parameters, as noted earlier, the cyclic prefix efficiency η_{CP} is $\sim$80% for all numbers of modes D considered. In Table V, values of T_{adapt} required for RLS are all at least a factor of ten below 25 μs for all values of D. Recalling (9), the relationship between the block length N_{FFT} and T_{adapt}, it is possible to increase N_{FFT} up to a factor of eight. For example, increasing N_{FFT} fourfold (to 2^{13}, 2^{13}, 2^{11}, 2^{11} for $D = 6$, 12, 20, 30) increases the cyclic prefix efficiency η_{CP} to $\sim$95% for

all values of D. In summary, the fast convergence of RLS can be exploited to improve cyclic prefix efficiency.

Computational complexity is closely related to transceiver power consumption, making it a key metric for comparing adaptive algorithms. Referring to Table IV, compared to LMS, RLS requires 2.0 to 2.7 times more complex multiplications/symbol and 1.5 to 2.1 times more complex additions/symbol. As shown in Table V, adaptation times can be up to ten times shorter for RLS than for LMS, so the total computational complexity to adapt to an unknown channel is lower for RLS than for LMS. If RLS uses a fourfold larger block length N_{FFT} to improve cyclic prefix efficiency η_{CP}, then the complexity slightly increases but still remains lower than for LMS. In summary, the fast convergence of RLS can be exploited to minimize the complexity required to adapt to an unknown channel.

In continuous tracking of a dynamic channel, the higher computational complexity of RLS becomes more of a concern. Complexity and power consumption might be minimized by performing the RLS updates (15) and (16) less frequently (since the duration of a block of N_{FFT} symbols is far less than 25 μs), or by employing LMS updates (14). For the first approach, there is also a possibility to avoid cyclic prefix overhead between adaptation intervals by employing overlap-save-based FDE using filter coefficients approximated from the FDE adapted using a cyclic prefix [47]. Evaluation of such channel tracking methods requires models for channel dynamics. A model for fibers with two mode groups ($D = 6$) has been proposed [30]. It is desirable to extend such models to larger D and to experimentally validate them.

We have chosen to employ a cyclic prefix instead of using block convolution for reasons stated in Sections II-C and III. As noted there, while this reduces throughput and transmit power efficiency by the factor η_{CP}, it significantly reduces the complexity of an adaptive FDE. This approach can be leveraged by inserting additional pilot symbols to aid in frame synchronization [59] or synchronization of carrier frequency [60] or phase [61], improving overall system performance and efficiency. The combination of cyclic prefix and pilot symbols is often called a "unique word".

In this study, we have neglected the impact of phase noise. There exist several candidate methods for phase noise compensation. Their implementation and integration with adaptive FDE are important subjects for further research. Since FDE can be more prone to degradation from phase error than TDE, it may be desirable to partially or fully compensate phase noise before the adaptive FDE, even if it is necessary to use phase estimates obtained after the adaptive FDE. MD equalization changes the phase noise statistics, an effect called "equalization-enhanced phase noise" [62], and commonly used symbol-by-symbol time-domain phase recovery methods, such as feedforward carrier recovery [63], will yield suboptimal performance unless modified to account for the modified phase statistics.

A technique popular in wireless systems with FDE is block-by-block phase estimation. With the aid of pilot symbols within a unique word [61], accurate phase estimation at the beginning of each block can be achieved [64]. The estimated phase is used to compensate the phase error of all symbols within the block prior to FDE. This approach can be implemented with low computational complexity without modifying the FDE adaptation algorithm. We have simulated this method assuming zero phase error at the beginning of each block and assuming the phase noise is a Wiener process [62] inside the block. We have observed negligible degradation in the asymptotic SER of adaptive FDE, assuming the parameters in Fig. 7 with $D = 12$ modes, provided the transmitter and local oscillator lasers have a beat linewidth not exceeding $\sim$65 kHz, which can be achieved using commercial tunable laser modules [65].

As [19], [28] and this paper demonstrate, the GD spread is a key factor in determining the computational complexity, adaptation speed, and cyclic prefix efficiency of an MDM system. In order to manage the GD spread, as explained in Section IV-A, we propose to use a GIGDC fiber with low uncoupled GD spread in conjunction with strong mode coupling described by a section length $L_{\mathrm{sec}} = 1$ km. This approach is considered in [20], where it is pointed out that manufacturing process variations may increase the uncoupled GD spread beyond its ideal value, and that splices between fiber sections may lead to section lengths $L_{\mathrm{sec}} \sim 5$ km. Thus, obtaining the low coupled GD spread assumed here may require intentionally perturbing the fiber in some way analogous to the "spinning" used to reduce PMD in SMF [66]. Fiber designs for reduced GD spread, including methods to enhance mode coupling without increasing loss and MDL, are important topics for future research.

Given the present uncertainty about mode coupling dynamics and coupled GD spreads achievable in long-haul fibers, the required adaptation time T_{adapt} and achievable impulse response durations N_{MD} assumed here, and the resulting FDE design parameters and performance and efficiency metrics, should be considered as illustrative examples more than precise determinations.

An alternate approach to GD management, which has been termed "GD compensation" reduces end-to-end GD spread by interconnecting two or more fiber types in which lower- and higher-order modes exhibit an opposite ordering of GDs [67]. This approach may be difficult to scale to several mode groups, where effective reduction of GD spread would require several fiber types with specific modal GD properties, as pointed out in [68]. Also, in certain cases, mode coupling may limit the reduction of GD spread, because in the strong-coupling regime, end-to-end rms GD spread depends only on the rms GD spreads of the individual segments [28], and not on the ordering of modal GDs.

VI. Conclusion

Long-haul MDM systems must typically use MIMO FDE to achieve sufficiently low computational complexity, in contrast to polarization multiplexing in SMF, where MIMO TDE has low complexity and is usually employed.

We have studied the LMS and RLS algorithms for adaptive MIMO FDE. Instead of using block convolution, we insert a cyclic prefix. While reducing throughput efficiency, this lowers the computational complexity of adaptation and may provide additional functionality, e.g., by facilitating carrier

TABLE VI
COMPUTATIONAL COMPLEXITIES OF ALGORITHM STEPS, WHICH ARE SUMMED TO OBTAIN RESULTS IN TABLE I

Equation	Computed term ($\mathbf{M}_i$ are $D\times D$ matrices, $\mathbf{V}_i$ are $D\times 1$ vectors and s_i are scalars)	Complex multiplications	Complex additions
(11)	$\mathbf{V_1} = \mathbf{W}_{MD}[k]\tilde{\mathbf{y}}[k]$	D^2	D^2-D
(14)	$\mathbf{V}_2 = \tilde{\mathbf{x}}[k]-\mathbf{W}_{MD}[k]\tilde{\mathbf{y}}[k]=\tilde{\mathbf{x}}[k]-\mathbf{V_1}$	-	D
	$\mathbf{V}_3{}^H = \tilde{\mathbf{y}}[k]^H\mu$	D	-
	$\mathbf{M_1} = (\tilde{\mathbf{x}}[k]-\mathbf{W}_{MD}[k]\tilde{\mathbf{y}}[k])\,\tilde{\mathbf{y}}[k]^H\mu = \mathbf{V}_2\,\mathbf{V}_3{}^H$	D^2	-
	$\mathbf{M}_2 = \mathbf{W}_{MD}[k]+(\tilde{\mathbf{x}}[k]-\mathbf{W}_{MD}[k]\tilde{\mathbf{y}}[k])\,\tilde{\mathbf{y}}[k] = \mathbf{W}_{MD}[k]+\mathbf{M_1}$	-	D^2
(15)	$\mathbf{M}_3 = \mathbf{R}[k]\kappa^{-1}$	D^2	-
	$\mathbf{V}_2 = \tilde{\mathbf{x}}[k]-\mathbf{W}_{MD}[k]\tilde{\mathbf{y}}[k]=\tilde{\mathbf{x}}[k]-\mathbf{V_1}$	-	D
	$\mathbf{V}_4{}^H = \tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1}) = \tilde{\mathbf{y}}[k]^H\mathbf{M}_3$	D^2	D^2-D
	$\mathbf{M}_4 = (\tilde{\mathbf{x}}[k]-\mathbf{W}_{MD}[k]\tilde{\mathbf{y}}[k])\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1}) = \mathbf{V}_2\mathbf{V}_4{}^H$	D^2	-
	$\mathbf{M}_5 = \mathbf{W}_{MD}[k]+(\tilde{\mathbf{x}}[k]-\mathbf{W}_{MD}[k]\tilde{\mathbf{y}}[k])\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1}) = \mathbf{W}_{MD}[k]+\mathbf{M}_4$	-	D^2
(16)	$\mathbf{V}_5 = (\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k] = \mathbf{M}_3\,\tilde{\mathbf{y}}[k]$	D^2	D^2-D
	$s_1 = \tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k] = \tilde{\mathbf{y}}[k]^H\mathbf{V}_5$	D	$D-1$
	$s_2 = 1+\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k] = 1+s_1$	-	1
	$\mathbf{V}_6 = \dfrac{(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k]}{1+\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k]} = \dfrac{\mathbf{V}_5}{s_2}$	D	-
	$\mathbf{M}_6 = \dfrac{(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k]\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1})}{1+\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k]} = \mathbf{V}_6\mathbf{V}_4{}^H$	D^2	-
	$\mathbf{M}_7 = (\mathbf{R}[k]\kappa^{-1}) - \dfrac{(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k]\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1})}{1+\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1})\tilde{\mathbf{y}}[k]} = \mathbf{M}_3-\mathbf{M}_6$	-	D^2

synchronization. We have investigated tradeoffs between computational complexity, cyclic prefix efficiency, adaptation speed and SER, and the impact of the system GD spread and FFT block length on these.

We show that using optimized FDE architectures, computational complexities increase sublinearly with the number of modes, in contrast to those using TDE. As compared to LMS, RLS achieves faster convergence, higher throughput efficiency, lower output SER, and greater tolerance to mode-dependent loss, but at the cost of somewhat higher complexity per FFT block. These attributes make RLS preferable for adapting to an unknown channel. For continuous tracking of a dynamic channel, either using RLS at periodic intervals or using LMS continuously might be preferable, depending on channel dynamics and system requirements.

Our paper illustrates that design of an MDM system to minimize GD spread is required to enable low computational complexity and fast adaptation. This is enabled here by strong mode coupling and by a new family of GIGDC fibers in which uncoupled GD spread decreases with an increasing number of modes.

APPENDIX A
COMPUTATIONAL COMPLEXITY

In this section, we derive the computational complexities per block for the update steps (14)–(16) given in Table I. We consider the optimal ordering of algebraic operations to minimize complexity. For example, multiplication of a matrix, vector and scalar can be done with lower complexity by first multiplying the matrix and the vector and then the scalar, or by first multiplying the vector and the scalar and then the matrix, as opposed to first multiplying the scalar and the matrix and then the vector. We also assume that any computed result can be reused at multiple steps without being recomputed. We assume that multiplication of an $M \times N$ matrix by an $N \times 1$ vector requires MN complex multiplications and $M\,(N-1)$ complex additions. We assume the DFT and inverse DFT operations are performed with radix-2 FFT algorithms, so each requires $N_{\mathrm{FFT}}\log_2(N_{\mathrm{FFT}})/2$ complex multiplications and $N_{\mathrm{FFT}}\log_2(N_{\mathrm{FFT}})$ complex additions for a vector of length N_{FFT}. The number of complex multiplications and additions for each computed term is shown in Table VI. The results in Table I are obtained by adding them.

REFERENCES

[1] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.
[2] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," *Opt. Expr.*, vol. 19, no. 17, pp. 16680–16696, Aug. 2011.
[3] E. Ip, E. Mateo, and W. Ting, "Reduced-complexity nonlinear compensation based on equivalent-span digital backpropagation," presented at the Int. Conf. Opt. Internet, Yokohoma, Japan, May 2012, Paper TuE.1.
[4] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," *J. Lightw. Technol.*, vol. 28, no. 6, pp. 939–951, Mar. 2010.

[5] T. Morioka, Y. Awaji, R. Ryf, P. J. Winzer, D. Richardson, and F. Poletti, "Enhancing optical communications with brand new fibers," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. s31–s42, Feb. 2012.

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

[7] S. Randel *et al.*, "Mode-multiplexed 6 × 20-GBd QPSK transmission over 1200-km DGD-compensated few mode fiber," presented at the Opt. Fiber Commun. Conf., Los Angeles, CA, USA, Mar. 2012, Paper PDP5 C.5.

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

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

[10] E. Ip and J. M. Kahn, "Fiber impairment compensation using coherent detection and digital signal processing," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 502–519, Feb. 2010.

[11] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Expr.*, vol. 16, no. 2, pp. 753–791, Jan. 2008.

[12] P. Krummrich, E.-D. Schmidt, W. Weiershausen, and A. Mattheus, "Field trial results on statistics of fast polarization changes in long haul WDM transmission systems," presented at the Optical Fiber Commun. Conf., Anaheim, CA, USA, 2005, Paper OThT6.

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

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

[15] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2033–2043, Aug. 2007.

[16] J. Renaudier *et al.*, "Linear fiber impairments mitigation of 40-Gbit/s polarization-multiplexed QPSK by digital processing in a coherent receiver," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 36–42, Jan. 2008.

[17] M. Kuschnerov *et al.*, "DSP for coherent single-carrier receivers," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3614–3622, Aug. 2009.

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

[19] S. Ö. Arik, D. Askarov, and J. M. Kahn, "Effect of mode coupling on signal processing complexity in mode-division multiplexing," *J. Lightw. Technol.*, vol. 31, no. 3, pp. 423–431, Feb. 2013.

[20] D. Peckham, Y. Sun, A. McCurdy, and R. Lingle, "Few-mode fiber technology for spatial multiplexing," in *Optical Fiber Telecommunications VI*, I. P. Kaminow, T. Li, and A. E. Willner, Eds. Amsterdam, The Netherlands: Elsevier, 2013.

[21] D. Askarov and J. M. Kahn, "Design of transmission fibers and doped fiber amplifiers for mode-division multiplexing," *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1945–1948, Nov. 2012.

[22] D. Marcuse, "The impulse response of an optical fiber with parabolic index profile," *Bell Syst. Tech. J.*, vol. 52, pp. 1169–1174, 1973.

[23] C. R. Fludger *et al.*, "Coherent equalization and POLMUX-RZ-DQPSK for robust 100-GE transmission," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 64–72, Jan. 2008.

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

[25] 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.

[26] S. Randel *et al.*, "6 × 56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6 × 6 MIMO equalization," *Opt. Exp.*, vol. 19, no. 17, pp. 16697–16707, Aug. 2011.

[27] B. Inan *et al.*, "DSP complexity of mode-division multiplexed receivers," *Opt. Expr.*, vol. 20, no. 9, pp. 10859–10869, Apr. 2011.

[28] K.-P. Ho and J. M. Kahn, "Mode coupling and its impact on spatially multiplexed systems," in *Optical Fiber Telecommunications VI*, I. P. Kaminow, T. Li, and A. E. Willner, Eds. Amsterdam: Elsevier, 2013.

[29] D. Rafique, S. Stylianos, and A. D. Ellis, "Impact of power allocation strategies in long-haul few-mode fiber transmission systems," *Opt. Exp.*, vol. 21, no. 9, pp. 10801–10809, May 2013.

[30] N. Bai and G. Li, "Adaptive frequency-domain equalization for mode-division multiplexed transmission," *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1918–1921, Nov. 2012.

[31] X. Chen *et al.*, "Characterization of dynamic evolution of channel matrix in two-mode fibers," presented at the Opt. Fiber Commun. Conf., Anaheim, CA, USA, Mar. 2013, Paper OM2 C.3.

[32] J. M. Wang and B. Daneshrad, "A comparative study of MIMO detection algorithms for wideband spatial multiplexing systems," in *Proc. IEEE Conf. Wireless Commun. Netw.*, Mar. 2005, vol. 1, pp. 408–413.

[33] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.

[34] J. Coon, S. Armour, M. Beach, and J. McGeehan, "Adaptive frequency domain equalization for single-carrier multiple-input multiple-output wireless transmissions," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3247–3256, Aug. 2005.

[35] M. V. Clark, "Adaptive frequency-domain equalization and diversity combining for broadband wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1385–95, Oct. 1998.

[36] S. Ö. Arik, J. M. Kahn, and K.-P. Ho, "MIMO signal processing in mode-division multiplexing," *IEEE Signal Process. Mag*, vol. 31, no. 2, pp. 25–34, Mar. 2014.

[37] R. N. Mahalati, D. Askarov, and J. M. Kahn, "Adaptive control of mode-dependent gain in multi-mode erbium-doped fiber amplifiers," presented at the IEEE Summer Topical Space Division Multiplexing Opt. Commun., Waikoloa, HI, USA, Jul. 2013.

[38] N. Benvenuto, R. Dinis, D. Falconer, and S. Tomasin, "Single-carrier modulation with nonlinear frequency-domain equalization: An idea whose time has come again," *Proc. IEEE*, vol. 98, no. 1, pp. 60–96, Jan. 2010.

[39] F. Pancaldi, G. Vitetta, R. Kalbasi, N. Al Dhahir, M. Uysal, and H. Mheidat, "Single-carrier frequency domain equalization," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 37–56, Sep. 2008.

[40] S. Randel *et al.*, "MIMO processing for space-division multiplexed transmission," presented at the Adv. Photon. Congr., Colorado Springs, Colorado, USA, 2012, Paper SpW3B.4.

[41] Y. Jung *et al.*, "First demonstration and detailed characterization of a multimode amplifier for space division multiplexed transmission systems," *Opt. Expr.*, vol. 19, no. 26, pp. B952–B957, Dec. 2011.

[42] D. Askarov and J. M. Kahn, "Design of multi-mode Erbium-doped fiber amplifiers for low mode-dependent gain," presented at the IEEE Summer Topical Spatial Multiplexing, Seattle, WA, USA, 2012, Paper WC2.2.

[43] K.-P. Ho and J. M. Kahn, "Mode-dependent loss and gain: statistics and effect on mode-division multiplexing," *Opt. Expr.*, vol. 19, no. 17, pp. 16612–16635, Aug. 2011.

[44] K.-P. Ho and J. M. Kahn, "Statistics of group delays in multimode fiber with strong mode coupling," *J. Lightw. Technol.*, vol. 29, no. 21, pp. 3119–3128, Nov. 2011.

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

[46] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[47] D. Falconer and S. L. Ariyavisitakul, "Broadband wireless using single carrier and frequency domain equalization," in *Int. Conf. Wireless Personal Multimedia Commun.*, Honolulu, HI, USA, Oct. 2002, pp. 27–36.

[48] D. J. F. Barros and J. M. Kahn, "Optimized dispersion compensation using orthogonal frequency-division multiplexing," *J. Lightw. Technol.*, vol. 26, no. 16, pp. 2889–2898, Aug. 2008.

[49] J. C. Geyer, C. Fludger, T. Duthel, C. Schulien, and B. Schaauss, "Efficient frequency domain chromatic dispersion compensation in a coherent Polmux QPSK-receiver," presented at the Opt. Fiber Commun. Conf., San Diego, CA, USA, 2010, Paper OWV5.

[50] G. Goldfarb and G. Li, "Chromatic dispersion compensation using digital IIR filtering with coherent detection," *IEEE Photon. Technol. Lett.*, vol. 19, no. 13, pp. 969–971, Jul. 2007.

[51] K. P. Ho, "Subband equaliser for chromatic dispersion of optical fibre," *Electron. Lett.*, vol. 45, no. 24, pp. 1224–1226, Nov. 2009.

[52] B. Li, Q. Wang, G. Lu, Y. Chang, and D. Yang, "Linear MMSE frequency domain equalization with colored noise," in *Proc. IEEE VTC*, Baltimore, USA, Sep. 2007, pp. 1152–1156.

[53] J. G. Proakis, *Digital Communications*, 3rd ed. New York, NY, USA: McGrawHill, 1995.

[54] D. Donlagic, "Opportunities to enhance multimode fiber links by application of overfilled launch," *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3526–3540, Nov. 2005.
[55] K. Okamoto and T. Okoshi, "Computer-aided synthesis of the optimum refractive-index profile for a multimode fiber," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-25, no. 3, pp. 213–221, Mar. 1977.
[56] T. Ishigure, H. Endo, K. Ohdoko, and Y. Koike, "High-bandwidth plastic optical fiber with W-refractive index profile," *IEEE Photon. Technol. Lett.*, vol. 16, no. 9, pp. 2081–2083, Sep. 2004.
[57] R. Feced, S. J. Savory, and A. Hadjifotiou, "Interaction between polarization mode dispersion and polarization dependent losses in optical communication links," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 20, no. 3, pp. 424–433, Mar. 2003.
[58] N. Cvijetic, E. Ip, N. Prasad, and M. Li, "Experimental frequency-domain channel matrix characterization for SDM-MIMO-OFDM systems," presented at the IEEE Summer Topical Space Division Multiplexing Opt. Commun., Waikoloa, HI, USA, Jul. 2013, Paper WC4.3.
[59] L. Deneire, B. Gyselinckx, and M. Engels, "Training sequence versus cyclic prefix-a new look on single carrier communication," *IEEE Commun. Lett.*, vol. 5, no. 7, pp. 292–294, Jul. 2001.
[60] M. Huemer, H. Witschnig, and J. Hausner, "Unique word based phase tracking algorithms for SC/FDE-systems," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2003, vol. 1, pp. 70–74.
[61] M. Asim, M. Ghogho, and D. McLemon, "Mitigation of phase noise in single carrier frequency domain equalization systems," in *Proc. Wireless Commun. Netw. Conf.*, Apr. 2012, pp. 920–924.
[62] K.-P. Ho and W. Shieh, "Equalization-enhanced phase noise in mode-division multiplexed Systems," *J. Lightw. Technol.*, vol. 31, no. 13, pp. 2237–2243, Jul. 2013.
[63] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2675–2692, Sep. 2007.
[64] M. Sabbaghian and D. Falconer, "Joint turbo frequency domain equalization and carrier synchronization," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 204–212, Jan. 2008.
[65] E. Ip, J. M. Kahn, D. Anthon, and J. Hutchins, "Linewidth measurements of MEMS-based tunable lasers for phase-locking applications," *IEEE Photon. Technol. Lett.*, vol. 17, no. 10, pp. 2029–2031, Oct. 2005.
[66] D. A. Nolan and M. J. Li, "Fiber spin-profile designs for producing fibers with low polarization mode dispersion," *Opt. Lett.*, vol. 23, no. 21, pp. 1659–1661, Nov. 1998.
[67] R. Ryf *et al.*, "Low-loss mode coupler for mode-multiplexed transmission in few-mode fiber," presented at the Proc. Opt. Fiber Commun. Conf., Los Angeles, CA, USA, Mar. 2012, Paper PDP5B.5.
[68] R. Ryf *et al.*, "12 × 12 MIMO transmission over 130-km few-mode fiber," presented at the Proc. Frontiers Opt. Conf., Rochester, NY, USA, Oct. 2012, Paper FW6 C.4.

Sercan Ö. Arık received the B.S. degree from Bilkent University, Ankara, Turkey, in 2011, and the M.S. degree from Stanford University in electrical engineering in 2013. He is currently working towards the Ph.D. degree at Stanford University. He worked at EPFL Signal Processing Labs (Lausanne, Switzerland) in summer 2010, at Google (Mountain View, CA, USA) in summer 2012 and at Mitsubishi Electric Research Labs (Cambridge, MA, USA) in summer 2013. His main current research interests include multi-mode optical communications; advanced modulation, coding and digital signal processing techniques for fiber-optic systems; and optical signal processing.

Daulet Askarov received the B.S. degree in applied mathematics and physics from Moscow Institute of Physics and Technology in 2009, and the M.S. degree in electrical engineering from Stanford University in 2011, where he is currently working toward the Ph.D. degree in electrical engineering. His current research interests include various topics in multi-mode fiber communications.

Joseph M. Kahn (M'90–SM'98–F'00) received the A.B., M.A., and Ph.D. degrees in physics from U.C. Berkeley in 1981, 1983, and 1986, respectively. From 1987–1990, he was at AT&T Bell Laboratories, Crawford Hill Laboratory, in Holmdel, NJ, USA. He demonstrated multi-Gbit/s coherent optical fiber transmission systems, setting world records for receiver sensitivity. From 1990–2003, he was with the Faculty of the Department of Electrical Engineering and Computer Sciences at U.C. Berkeley, performing research on optical and wireless communications. Since 2003, he has been a Professor of Electrical Engineering at Stanford University, where he heads the Optical Communications Group. His current research interests include fiber-based imaging, spatial multiplexing, rate-adaptive and spectrally efficient modulation and coding methods, coherent detection and associated digital signal processing algorithms, digital compensation of fiber nonlinearity, and free-space systems. He received the National Science Foundation Presidential Young Investigator Award in 1991. From 1993–2000, he served as a technical editor of IEEE *Personal Communications Magazine.* Since 2009, he has been an Associate Editor of IEEE/OSA *Journal of Optical Communications and Networking.* In 2000, he helped found StrataLight Communications, where he served as Chief Scientist from 2000–2003. StrataLight was acquired by Opnext, Inc., in 2009.