Sercan Ö. Arık, Joseph M. Kahn, and Keang-Po Ho

MIMO Signal Processing for Mode-Division Multiplexing

An overview of channel models and signal processing architectures

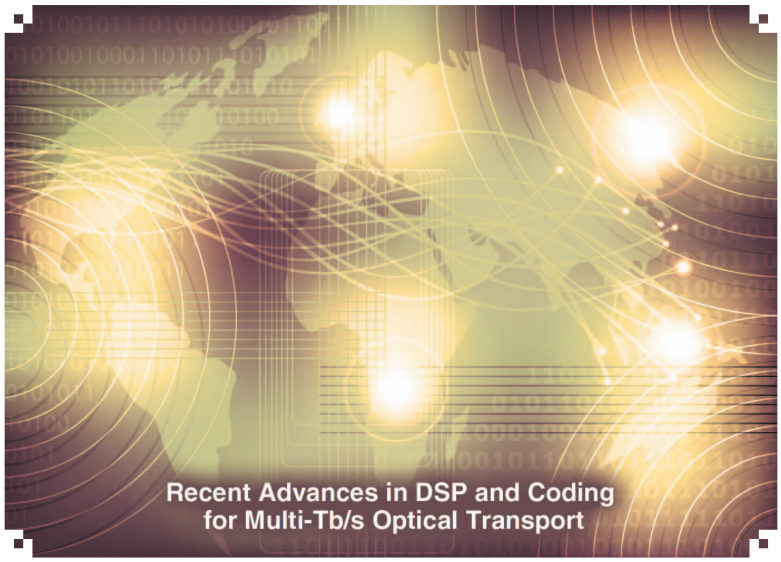


IMAGE LICENSED BY INGRAM PUBLISHING /JANMIKS

We present the fundamentals of multiple-input, multiple-output (MIMO) signal processing for mode-division multiplexing (MDM) in multimode fiber (MMF). As an introduction, we review current long-haul optical transmission systems and how continued traffic growth motivates study of new methods to increase transmission capacity per fiber. We describe the key characteristics of MIMO channels in MMF, contrasting these with wireless MIMO channels. We review MMF channel models, the statistics derived from them, and their implications for MDM system performance and complexity. We show that optimizing performance and complexity requires management of channel parameters—particularly group delay (GD) spread and mode-dependent loss and gain—by design of transmission fibers and optical amplifiers, and by control of mode coupling along the link. We describe a family of fibers optimized for low GD spread, which decreases with an increasing number of modes. We compare the performance and complexity of candidate MIMO signal processing architectures in a representative long-haul system design, and show that programmable frequency-domain equalization (FDE) of chromatic dispersion (CD) and adaptive FDE of modal dispersion (MD) is an attractive combination. We review two major algorithms for adaptive FDE of MD—least mean

Digital Object Identifier 10.1109/MSP.2013.2290804
Date of publication: 12 February 2014

squares (LMS) and recursive least squares (RLS)—and analyze their complexity, throughput efficiency, and convergence time. We demonstrate that, with careful physical link design and judicious choice of signal processing architectures, it is possible to overcome MIMO signal processing challenges in MDM systems.

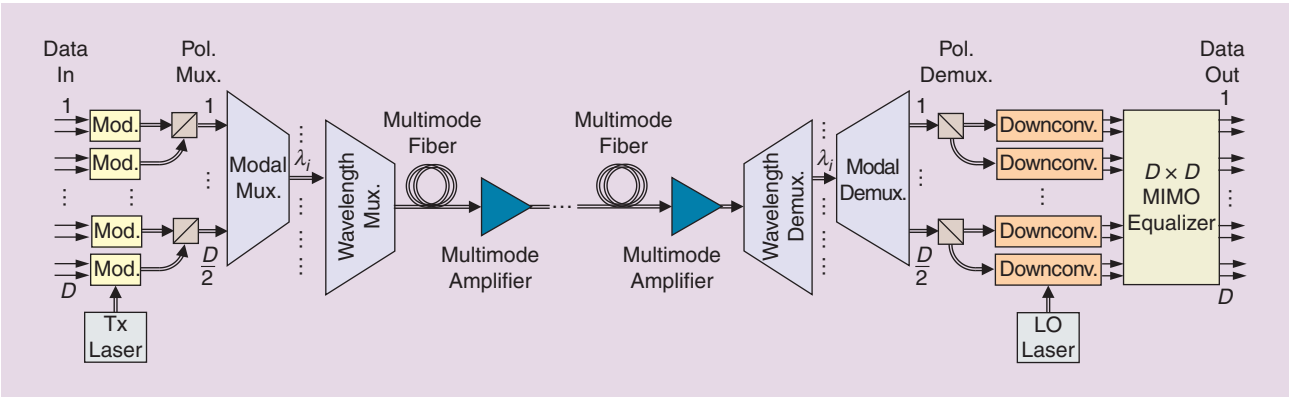
INTRODUCTION

Spurred by high-definition video streaming, multimedia file sharing, cloud computing, mobile networking, online gaming, and other information technologies, worldwide data traffic is growing at a rate estimated to exceed 50% annually [1]. Long-haul optical fiber networks form the backbone of the Internet, and scaling to higher throughput and lower cost per bit is essential to ensure continued growth of information technologies. For four decades, optical networks have exploited significant progress in physical technologies, including low-loss single-mode fibers (SMFs), high-performance lasers and modulators, and low-noise optical amplifiers. In parallel, communication methods have evolved to multiplex information in various physical dimensions in SMFs, including time, frequency, quadrature phase and polarization, a trend hastened in recent years by coherent optical receivers using large-scale digital circuits for signal processing and error-correction decoding. The transmission capacity per fiber, after decades of exponential growth sustained by these technologies [2], is now approaching fundamental information-theoretic limits imposed by optical amplifier noise and by the nonlinear response of the silica fiber medium [3]. Digital signal processing methods can mitigate nonlinear effects [4], enabling denser constellations and increasing transmission capacity, but at the cost of high complexity and power consumption.

New types of fibers, which provide more spatial dimensions, provide intriguing options for increasing transmission capacity per fiber. Multicore fibers provide one option [2]. By incorporating multiple cores in a single strand of glass, they provide a proportional increase in the number of spatial dimensions. The plurality of modes propagating in MMF provides another option for increasing the number of spatial dimensions [2]. A waveguide mode is defined as a pattern of the optical electric field that propagates without changing, apart from an amplitude change and phase shift. Different waveguide modes are mutually orthogonal, so they provide independent spatial dimensions for data transmission. Although propagation over long distances causes coupling between signals in different modes, received signals can be separated by MIMO processing. Throughout this article, we use D to denote the total number of modes, including spatial and polarization dimensions. A "single-mode" fiber allows propagation of one spatial mode in two polarizations, a total of $D = 2$ modes. A 2×2 MIMO transmission in SMF is also known as polarization-division multiplexing (PDM). By increasing the core diameter and adjusting the refractive index profile, an MMF with circular cross section can be designed to support $D = 6, 10, 12, 16, 20, 24, 30, \ldots$ modes. Fibers supporting these small values of D are sometimes called "few-mode fibers" to distinguish them from the MMFs used in short-range data communications, which support up to hundreds of modes.

WITH CAREFUL PHYSICAL LINK DESIGN AND JUDICIOUS CHOICE OF SIGNAL PROCESSING ARCHITECTURES, IT IS POSSIBLE TO OVERCOME MIMO SIGNAL PROCESSING CHALLENGES IN MDM SYSTEMS.

Figure 1 shows the general structure of a long-haul MDM transmission system. At the transmitter, at each wavelength, D signals are modulated with data streams and mapped into two polarizations and $D/2$ spatial modes, a fixed transmit basis comprising D orthogonal modes. Signals from up to about 100 transmitters at different wavelengths are combined by a wavelength multiplexer, as in a conventional SMF system. As signals propagate through multiple spans of fiber, loss is compensated by optical amplifiers designed to amplify multiple modes at multiple wavelengths. At the receiving end, different wavelengths are demultiplexed. The signal at each wavelength is separated into $D/2$ spatial modes and two polarizations, i.e., projected onto a fixed receive basis of D modes. The signal in each mode is mixed with a local oscillator, downconverting it to baseband. After sampling, digital signal processing implements several essential functions, which



[FIG1] A long-haul optical transmission system using mode-division multiplexing.

[TABLE 1] A COMPARISON OF MIMO CHANNELS IN WIRELESS SYSTEMS AND IN MDM OPTICAL SYSTEMS.

ATTRIBUTE	WIRELESS	OPTICAL
CHANNEL INPUTS AND OUTPUTS	SIGNALS AT ANTENNAS	COMPLEX AMPLITUDES IN MODES
NUMBER OF INPUTS OR OUTPUTS	2, 3, 4, 5, 6,…	2, 6, 10, 12, 16, 20, 24, 30,…
CARRIER FREQUENCY	0.8–6 GHz (MOST PHONE AND Wi-Fi BANDS)	185–196 THz (OPTICAL AMPLIFIER BANDS)
SYMBOL RATE	10 s OF Mbaud	10 s OF Gbaud
ADDITIVE NOISE	THERMAL NOISE (CLASSICAL)	SPONTANEOUS EMISSION (QUANTUM)
INTERFERENCE AND DISTORTION	COCHANNEL INTERFERENCE (LINEAR)	INTRACHANNEL DISTORTION (NONLINEAR) INTERCHANNEL INTERFERENCE (NONLINEAR)
DISPERSION	MULTIPATH PROPAGATION	FREQUENCY-DEPENDENT GROUP INDEX (CD) MODE-DEPENDENT GROUP INDEX (MD)
DELAY SPREAD (rms)	10 s OF ns (INDOORS) 100 s OF ns (OUTDOORS)	100 s OF ns (LONG HAUL)
FADING	MULTIPATH PROPAGATION	OPTICAL AMPLIFIER MODE-DEPENDENT GAIN
AMPLITUDE VARIATIONS IN CHANNEL MATRIX	LARGE (e.g., I.I.D. RAYLEIGH)	SMALL (e.g., NEAR-UNITARY)
ORIGIN OF CHANNEL VARIATIONS	TERMINAL MOTION (DOPPLER)	FIBER MOTION (MODE COUPLING)
CHANNEL VARIATION TIME SCALE	~1 ms (DRIVING OR WALKING SPEEDS)	10 TO 100 s OF μs (VIBRATION OR IMPACT)
DIVERSITY	MAY REQUIRE SPACE-TIME CODING	INHERENT FREQUENCY DIVERSITY

include automatic gain control, timing recovery, carrier recovery, and MIMO equalization that compensates for CD, MD, and mode coupling, separating the D multiplexed data signals [5].

Making long-haul MDM systems a practical reality requires development of new fibers and other optical components supporting multiple spatial modes, including modal (de)multiplexers, wavelength (de) multiplexers, wavelength-selective switches, and inline optical amplifiers. Implementing transceivers supporting multiple spatial modes while reducing the cost, size, and power consumption per information bit will require extensive photonic and electronic integration. Likewise, it will require efficient, high-performance MIMO signal processing architectures, which are the subject of this article.

MAKING LONG-HAUL MDM SYSTEMS A PRACTICAL REALITY REQUIRES DEVELOPMENT OF NEW FIBERS AND OTHER OPTICAL COMPONENTS SUPPORTING MULTIPLE SPATIAL MODES.

COMPARISON OF WIRELESS AND OPTICAL MIMO TRANSMISSION

Over the past 15 years, MIMO wireless transmission has progressed from its information-theoretic foundations [6] to recent commercialization. MDM, as an optical form of MIMO transmission, shares some similarities with wireless MIMO transmission. But optical and wireless MIMO channels exhibit some fundamental differences. A comparison of these two types of MIMO channels is given in Table 1.

Wireless systems use lower carrier frequencies, where bandwidth is limited and regulated. Optical systems use far higher carrier frequencies, where bandwidth is limited mainly by the gain spectra of optical amplifiers. In wireless, the dominant noise is thermal, since the carrier photon energy is far less than thermal energy, while in dense user environments, linear cochannel interference from other users can become the dominant "noise." In optical systems, where the carrier photon energy is far greater than thermal energy, the dominant noise is spontaneous emission from inline optical amplifiers. In glass fibers, the refractive index is weakly dependent on the local light intensity. This so-called Kerr nonlinearity causes self-phase modulation, a form of intrachannel distortion, and also cross-phase modulation and four-wave mixing, which are forms of adjacent-channel interference. In current systems, these nonlinear effects are treated as "noise." It is possible to mitigate them partially by digital backpropagation [4], which would allow some increase in spectral efficiency, but at the cost of increased signal processing complexity, which precludes its implementation in current hardware technologies.

In wireless MIMO systems, multipath propagation causes linear distortion, with a delay spread depending mainly on the environment. Multipath also causes fading, which can cause the signal energy in a narrow band to vary by tens of decibels. These wireless channel properties vary randomly at rates depending on the speed of transceiver or scatterer motion. In a narrowband MIMO system, fading causes the instantaneous capacity to fluctuate, potentially causing an outage, and if a system does not provide sufficient frequency or time diversity, then space-time coding may be required to achieve reliable operation [6].

In optical MIMO systems, linear distortion arises from two mechanisms. First, the group refractive index of a mode is frequency dependent, an effect called CD that is similar in SMF and MMF. At frequencies employed in long-haul systems, CD arises from the inherent material dispersion of glass and from frequency-dependent confinement of the mode in the waveguide, an opposing contribution that is typically smaller in magnitude. Second, the group refractive index is slightly different for different modes. In SMF, with only $D = 2$ polarization modes, this is called polarization-mode dispersion (PMD) [7] and causes a delay spread of

only 10s of ps. In MMF, the effect is called MD and can cause a delay spread up to 100s of ns. In optical MIMO systems, fading arises because different modes may experience different gains in optical amplifiers (or different losses in fibers), effects collectively called mode-dependent loss (MDL) [8]. In SMF $D = 2$, this effect is called polarization-dependent loss (PDL). If managed carefully, root-mean-square (rms) MDL fluctuations are expected to be smaller than 10 dB, so the channel matrix remains near-unitary. Both MD and MDL may vary randomly over time and frequency because of random coupling between modes caused by perturbations of the fiber, including index imperfections and bends [9]. In response to mechanical perturbations of the fiber, optical MIMO channels are expected to fluctuate on the time scale of 10 s–100 s of μs [10], [11]. In optical MIMO systems, the product of delay spread and symbol rate far exceeds unity, yielding efficient frequency diversity that reduces capacity fluctuations [12]. As a result, all D spatial dimensions can be exploited for multiplexing, and none need be sacrificed for diversity.

A final comparison concerns the choice of single- versus multicarrier modulation. The latter is often implemented as orthogonal frequency-division multiplexing (OFDM). Comparison between the two approaches is a perennial topic in wireless research and has motivated study of multicarrier techniques for optical PDM and MDM [13]. In wireless communications, when single- and multicarrier systems are optimally designed, they achieve similar performance and complexity. By contrast, in long-haul optical systems, single-carrier modulation is less impacted by fiber nonlinearity than multicarrier modulation [14], and single-carrier has been used in all commercial systems to date. Nevertheless, the single-carrier FDE methods described below can be adapted to OFDM with only minor modifications.

MDM CHANNEL MODEL AND STATISTICS

In this section, we describe the modeling of MDM channels in MMF, emphasizing the effect of mode coupling on the channel GD spread and fading statistics. We provide an MDM system example to illustrate how the GD spread affects MIMO signal processing complexity and performance.

MATRIX PROPAGATION MODEL

A signal at frequency Ω propagating along the z-axis in a MMF can be represented as an electric field vector $\mathbf{E}(x, y, z, \Omega) = \sum_{i=1}^{D} A_i(z, \Omega)\mathbf{E}_i(x, y, \Omega)$, where the $\mathbf{E}_i(x, y, \Omega)$, $i = 1, \ldots, D$, are the orthonormal waveguide modes of an ideal unperturbed fiber. Since these modes are fixed, we can represent the propagating signal by a $D \times 1$ vector of complex amplitudes $\mathbf{A}(z, \Omega) = [A_1(z, \Omega) \ \cdots \ A_D(z, \Omega)]^T$, which we write as $\mathbf{A}(\Omega)$. Ignoring noise, linear propagation through the fiber can be described by

$$\mathbf{A}^{(\mathrm{out})}(\Omega) = \mathbf{M}(\Omega)\mathbf{A}^{(\mathrm{in})}(\Omega), \tag{1}$$

where $\mathbf{A}^{(\mathrm{in})}(\Omega)$ and $\mathbf{A}^{(\mathrm{out})}(\Omega)$ represent the input and output and $\mathbf{M}(\Omega)$ is a propagation operator described by a $D \times D$ matrix. In SMF ($D = 2$), the $\mathbf{A}(z, \Omega)$ and $\mathbf{M}(\Omega)$ are the Jones vector and Jones matrix [7].

Because of perturbations inducing mode coupling, signals evolve randomly as they propagate through a fiber, and their complex envelopes remain correlated over a characteristic correlation length. A fiber can be modeled as a concatenation of multiple sections [9], each of length roughly equal to the correlation length. This is a generalization of the multisection model used in modeling PMD and PDL in SMF [15]. As shown in Figure 2, a long-haul 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 $L_{\mathrm{tot}} = K_{\mathrm{amp}} L_{\mathrm{amp}} = K_{\mathrm{amp}} K_{\mathrm{sec}} L_{\mathrm{sec}}$.

The propagation operator can be represented as [9]

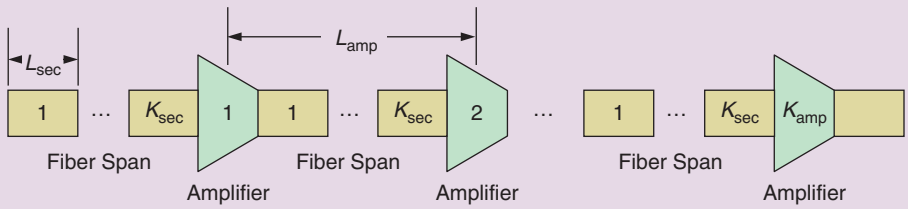
$$\mathbf{M}_{\mathrm{tot}}(\Omega) = \exp\left(-\frac{j}{2}\Omega^2 \bar{\beta}_2 L_{\mathrm{tot}}\right) \cdot \mathbf{M}(\Omega). \tag{2}$$

The exponential factor represents mode-averaged propagation, where $\bar{\beta}_2$ represents mode-averaged CD per unit length (for simplicity, we neglect the mode-averaged GD). The matrix $\mathbf{M}(\Omega)$ represents mode-dependent effects, including MDL, MD, and crosstalk caused by mode coupling. It can be written as a product over the K_{amp} spans

$$\mathbf{M}(\Omega) = \prod_{k=1}^{K_{\mathrm{amp}}} \mathbf{M}^{(k)}(\Omega), \tag{3}$$

where $\mathbf{M}^{(k)}(\Omega)$ represents propagation in the kth span. Assuming MDL in the transmission fibers is negligible compared to that in the amplifiers [5], it can be written

$$\mathbf{M}^{(k)}(\Omega) = \mathrm{diag}\left[\exp\left(\frac{g_1^{(k)}}{2}\right) \ \ldots \ \exp\left(\frac{g_D^{(k)}}{2}\right)\right] \cdot \prod_{l=1}^{K_{\mathrm{sec}}} \mathbf{V}^{(k,l)} \mathbf{\Lambda}(\Omega) \mathbf{U}^{(k,l)H}. \tag{4}$$



[FIG2] The multisection model of a long-haul MDM system.

The first factor in $\mathbf{M}^{(k)}(\Omega)$ describes uncoupled MDL in the kth amplifier. The $g_i^{(k)}$, $i = 1, \ldots, D$, are the uncoupled modal gains measured in log power gain units (proportional to decibels). They satisfy $\sum_{i=1}^{D} g_i^{(k)} = 0$ and have rms spread σ_g (assumed to be the same for all amplifiers). The second factor in $\mathbf{M}^{(k)}(\Omega)$ describes MD and mode coupling in the kth span and is a product over the K_{sec} sections. The matrix $\mathbf{\Lambda}(\Omega)$ describes uncoupled MD in each section (assumed to be the same for all sections)

$$\mathbf{\Lambda}(\Omega) = \text{diag}[\exp(-j\Omega\tau_1) \ \ldots \ \exp(-j\Omega\tau_D)], \tag{5}$$

where the τ_i, $i = 1, \ldots, D$, are the uncoupled modal GDs, which satisfy $\sum_{i=1}^{D} \tau_i = 0$ and have rms spread σ_τ. The matrices $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are frequency-independent unitary matrices representing the random mode coupling in the lth section of the kth span. The symbol ^{H} denotes Hermitian conjugate.

Mode coupling in fibers may be induced by random or intentional index perturbations, bends or stresses. Many random perturbations have low-pass longitudinal power spectra, so they strongly couple modes having nearly equal propagation constants, but weakly couple modes having highly unequal propagation constants [9]. When mode coupling occurs only between nearly degenerate modes in the same mode group, which is the weak-coupling regime, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are block unitary matrices. When mode coupling occurs with approximately equal strength between all modes, $\mathbf{V}^{(k,l)}$ and $\mathbf{U}^{(k,l)}$ are fully random unitary matrices. Further assuming the correlation length is short relative to the system length, so the number of independent sections is large, this corresponds to the strong-coupling regime. In this regime, as shown below, the GD spread and the MDL spread are reduced, and the statistics of MD and MDL approach limiting distributions. Strong mode coupling is highly beneficial in long-haul MDM systems.

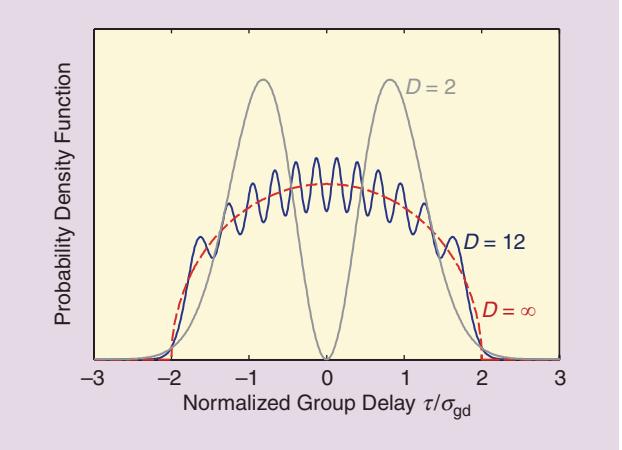
MODAL DISPERSION

Having described the channel model, we discuss the effect of mode coupling on the statistics of the modal GDs, which governs MIMO equalizer complexity. We ignore MDL for simplicity. Given the operator $\mathbf{M}(\Omega)$ describing MD and mode coupling, a GD operator is defined as $\mathbf{G} = j(\partial\mathbf{M}(\Omega)/\partial\Omega)\mathbf{M}(\Omega)^H$ [16], whose eigenvalues are the coupled GDs $(\tau_1^{(t)}, \ldots, \tau_D^{(t)})$, ordered as $\tau_1^{(t)} \le \tau_2^{(t)} \le \cdots \le \tau_D^{(t)}$.

Assume that in the absence of coupling, the fiber has peak-to-peak (p-p) and rms GD spread per unit length $\Delta\beta_{1,\text{pp}}$ and $\Delta\beta_{1,\text{rms}}$. Then in each section, the uncoupled GDs τ_i, $i = 1, \ldots, D$ in (5) have p-p spread $\Delta\beta_{1,\text{pp}} L_{\text{sec}}$ and rms spread $\sigma_\tau = \Delta\beta_{1,\text{rms}} L_{\text{sec}}$.

In the weak-coupling regime, the p-p coupled GD spread of the system is $\tau_D^{(t)} - \tau_1^{(t)} = \Delta\beta_{1,\text{pp}} K_{\text{sec}} L_{\text{sec}} = \Delta\beta_{1,\text{pp}} L_{\text{tot}}$, a deterministic quantity that scales linearly with the total number of sections K_{tot} or the total fiber length L_{tot}.

In the strong-coupling regime, which assumes all modes couple equally and $K_{\text{tot}} >> 1$, the coupled GDs $\left(\tau_1^{(t)}, \ldots, \tau_D^{(t)}\right)$ are random variables. The rms coupled GD spread of the system is $\sigma_{\text{gd}} = \sqrt{K_{\text{tot}}} \cdot \sigma_\tau = \Delta\beta_{1,\text{rms}} \sqrt{L_{\text{sec}} L_{\text{tot}}}$, which scales with the square root of K_{tot} or L_{tot} [9]. By the central limit theorem, the GD operator becomes a zero-trace Gaussian unitary ensemble. The statistics of the coupled GDs are given by the statistics of the eigenvalues of the ensemble, and depend only on D and σ_{gd} [9]. Figure 3 shows the probability density function (pdf) of the coupled GDs, normalized by σ_{gd}, for several values of D. When $D = 2$, the pdf is Maxwellian, which is known from the study of PMD [17]. For any D, the pdf peaks at D values of GD. In the limit $D \to \infty$, the pdf approaches a semicircle with a finite support $4\sigma_{\text{gd}}$. The presence of MDL is expected to change the GD statistics slightly, as noted for $D = 2$ [18]. The p-p coupled GD spread $\left(\tau_D^{(t)} - \tau_1^{(t)}\right)$ determines the temporal memory required in an equalizer for compensating MD [19], [20], as discussed below.



[FIG3] Pdfs of normalized coupled GDs for $D = 2, 12$, and ∞ modes.

MODE-DEPENDENT GAIN AND LOSS

We now discuss the impact of mode coupling on the statistics of MDL, which is the fading distribution governing the channel capacity of MDM systems [8]. Given the operator $\mathbf{M}(\Omega)$ including MDL, MD and mode coupling, an MDL operator is defined as $\mathbf{M}(\Omega)\mathbf{M}(\Omega)^H$. Its eigenvalues are the system's coupled MDL values $(g_1^{(t)}(\Omega), \ldots, g_D^{(t)}(\Omega))$, ordered as $g_1^{(t)}(\Omega) \ge g_2^{(t)}(\Omega) \ge \cdots \ge g_D^{(t)}(\Omega)$, measured in log power gain units. These are the logarithms of the spatial subchannel gains that determine MIMO channel capacity [8].

In the weak-coupling regime, the system's p-p MDL spread $g_1^{(t)}(\Omega) - g_D^{(t)}(\Omega)$ is a deterministic quantity that scales linearly with the number of amplifiers K_{amp}.

In the strong-coupling regime, which assumes all modes couple equally and the number of MDL sources (amplifiers) is large, $K_{\text{amp}} >> 1$, the system's coupled gains $(g_1^{(t)}(\Omega), \ldots, g_D^{(t)}(\Omega))$ are random variables. The statistics of coupled MDL depend only on the number of modes D and on the rms accumulated MDL $\xi = \sqrt{K_{\text{amp}}} \cdot \sigma_g$ [8], which scales with the square root of K_{amp}. Equivalently, these statistics depend on the rms overall MDL $\sigma_{\text{mdl}} = \xi\sqrt{1 + \xi^2/(12(1 - D^{-2}))}$ [18]. In the low-to-moderate MDL range that is of practical interest, the logarithm of the MDL operator is approximately a zero-trace Gaussian unitary ensemble

[8], so the statistics of MDL, normalized by σ_{mdl}, are described by the same distributions as the normalized GDs, shown in Figure 3. Although MDL can potentially make optical amplifier noise spatially nonwhite, this noise becomes spatially white in the strong-coupling regime with a large number of noise sources $K_{\mathrm{amp}} >> 1$ [8].

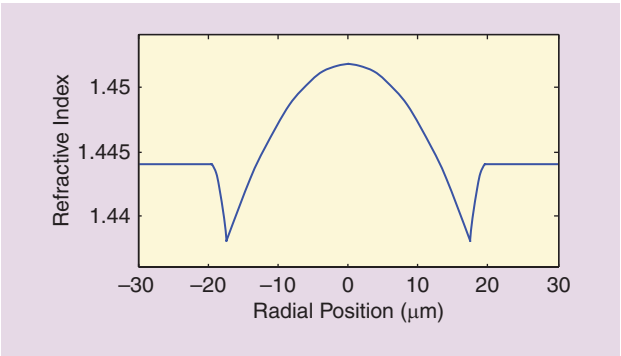
MANAGING AN MDM SYSTEM'S END-TO-END GD SPREAD IS CRUCIAL IN CONTROLLING THE COMPLEXITY AND PERFORMANCE OF MIMO SIGNAL PROCESSING.

In long-haul MDM systems, round-trip propagation delay may be tens of ms, and reliable channel state information (CSI) is not available at the transmitter. At any one frequency Ω, given a realization of the subchannel gains $(g_1^{(t)}(\Omega), \ldots, g_D^{(t)}(\Omega))$, the capacity (per unit frequency) is $C = \sum_{i=1}^{D} \log_2\left(1 + \rho \exp(g_i^{(t)}(\Omega))\right)$, where ρ is the mode-averaged transmitted signal power divided by the mode-averaged received noise power [8]. As the rms accumulated MDL increases, the average capacity decreases, and the variance of capacity increases, increasing the probability of outage [8]. Fortunately, the subchannel gains $(g_1^{(t)}(\Omega), \ldots, g_D^{(t)}(\Omega))$ vary over frequency, with a correlation bandwidth of approximately $1/\sigma_{\mathrm{gd}}$. As long as the symbol rate R_s is high enough that $R_s \sigma_{\mathrm{gd}} >> 1$ (a condition easily satisfied in practice), efficient frequency diversity reduces the outage probability and the outage capacity approaches the average capacity [12].

DELAY SPREAD: MANAGEMENT AND IMPACT

Managing an MDM system's end-to-end GD spread is crucial in controlling the complexity and performance of MIMO signal processing. Here, we review strategies for GD spread management and present an MDM system example to help illustrate how the GD spread affects MIMO signal processing.

Several early MDM experiments, e.g., [21], used fibers supporting two mode groups $(D = 6)$, where a low GD spread can be realized by choosing a core radius at which the GD-versus-radius curves for the two mode groups intersect [22]. This approach does not scale easily beyond two mode groups, since the curves for different pairs of modes intersect at different radii. An alternate approach minimizes GD spread by interconnecting different fiber types in which lower- and higher-order modes exhibit an opposite ordering of GDs [23]. However, this approach may be difficult to scale to several mode groups, since specific lengths of several fiber types with specific GD properties may be required. Also, mode coupling may affect the GD compensation obtained. Hence, we adopt the approach described in [22], using fibers with low uncoupled GD spread and relying on strong mode coupling, induced by splices or other perturbations, to further reduce the GD spread [9].



[FIG4] A fiber design with graded-index core and graded depressed cladding for $D = 12$ modes.

We consider a family of fibers with graded-index core and graded depressed cladding (GIGDC), which have the desirable properties that the rms uncoupled GD spread $\Delta\beta_{1,\mathrm{rms}}$ decreases with an increasing number of modes D [24]. The index profile for $D = 12$ is shown in Figure 4. Fabrication techniques for these kinds of fibers are discussed in [25] and [26].

To compensate for CD, a digital equalizer must have a duration, measured in samples, given by [27]

$$N_{\mathrm{CD}} = \left\lceil 2\pi |\bar{\beta}_2| L_{\mathrm{tot}} (r_{\mathrm{os}} R_s)^2 \right\rceil, \tag{6}$$

where r_{os} is the oversampling rate, and $\lceil x \rceil$ denotes the ceiling function. Note that N_{CD} scales linearly with the fiber length L_{tot} and quadratically with the symbol rate R_s.

To compensate for MD and modal crosstalk, a MIMO equalizer must have a duration sufficient to span the p-p coupled GD spread $\left(\tau_D^{(t)} - \tau_1^{(t)}\right)$. Measured in samples, the required duration is $N_{\mathrm{MD}} = \lceil (\tau_D^{(t)} - \tau_1^{(t)}) r_{\mathrm{os}} R_s \rceil$, which scales linearly with R_s. In the weak-coupling regime, the p-p coupled GD spread $\left(\tau_D^{(t)} - \tau_1^{(t)}\right)$ is deterministic, as noted above, and the required equalizer duration, measured in samples, is

$$N_{\mathrm{MD}} = \lceil \Delta\beta_{1,\mathrm{pp}} K L_{\mathrm{sec}} r_{\mathrm{os}} R_s \rceil = \lceil \Delta\beta_{1,\mathrm{pp}} L_{\mathrm{tot}} r_{\mathrm{os}} R_s \rceil, \tag{7}$$

which scales linearly with the number of sections K_{tot} or the fiber length L_{tot}. In the strong-coupling regime, $\left(\tau_D^{(t)} - \tau_1^{(t)}\right)$ is random. As in [19] and [20], we define $u_D(p)$ such that an equalizer duration $\sigma_{\mathrm{gd}} u_D(p)$ is longer than $\left(\tau_D^{(t)} - \tau_1^{(t)}\right)$ with probability $1 - p$. For typical values of D and $p \sim 10^{-4}$ to 10^{-6}, $u_D(p) \sim 4$ to 5, as expected from Figure 3. The required equalizer duration, measured in samples, is

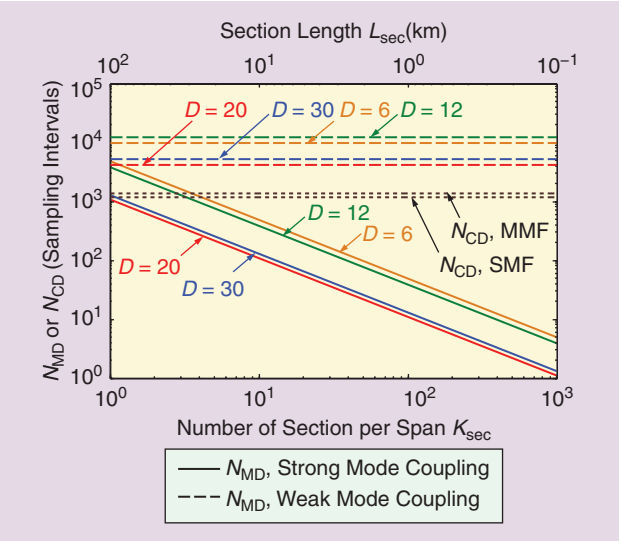
$$N_{\mathrm{MD}} = \left\lceil \sqrt{K_{\mathrm{tot}}}\, \sigma_\tau u_D(p) r_{\mathrm{os}} R_s \right\rceil = \left\lceil \Delta\beta_{1,\mathrm{rms}} \sqrt{L_{\mathrm{sec}} L_{\mathrm{tot}}}\, u_D(p) r_{\mathrm{os}} R_s \right\rceil, \tag{8}$$

which scales with the square-root of K_{tot} or L_{tot}.

To compare the equalizer memory length requirements (6)–(8), we consider a long-haul transmission system with 20 fiber spans, each of length $L_{\mathrm{amp}} = 100$ km, assuming $R_s = 32$ Gbaud, $r_{\mathrm{os}} = 2$ and $p = 10^{-5}$. Figure 5 shows values of N_{CD} and N_{MD} for GIGDC fibers supporting $D = 6$, 12, 20, or 30 modes, as a function of K_{sec} or $L_{\mathrm{sec}} = L_{\mathrm{amp}}/K_{\mathrm{sec}}$, comparing

these to N_{CD} for standard SMF. In the weak-coupling regime, values of N_{MD} for the GIGDC fibers are up to an order of magnitude larger than the values of N_{CD}. In the strong-coupling regime, as K_{sec} increases from 10^0 to 10^3 and L_{sec} decreases from 10^2 to 10^{-1} km, N_{MD} becomes up to two or three orders of magnitude smaller than the values of N_{CD}. As shown in [19] and discussed below, to counter the complexity associated with the $D \times D$ matrix structure of the MIMO equalizer, it is desirable to have N_{MD} at least an order of magnitude smaller than N_{CD}, which requires strong-mode coupling, even with these optimized fibers.

Finally, for comparison purposes, we evaluate the memory length of PMD in standard SMF, denoted by N_{PMD}. Assuming $\Delta\beta_{1,\mathrm{rms}}\sqrt{L_{sec}} = 0.1\ \mathrm{ps}/\sqrt{\mathrm{km}}$ and $u_D(p) = 5$, (8) yields $N_{PMD} = 2$, which is far smaller than N_{CD}.



[FIG5] The memory length (in sampling intervals) required to compensate CD or MD, versus number of sections per span K_{sec} or section length L_{sec}, for a long-haul transmission with 20 spans of GIGDC fiber, each 100-km long, with symbol rate $R_s = 32$ Gbaud and oversampling ratio $r_{os} = 2$.

RECEIVER SIGNAL PROCESSING

ARCHITECTURES

Receivers for PDM in SMF $(D = 2)$ employ homodyne (or intradyne) downconversion and high-speed analog-to-digital conversion, followed by fully digital processing [14], [27]. Digital equalization compensates for linear channel effects, including CD, PMD, and polarization crosstalk. Other functions implemented digitally may include automatic gain control, timing recovery, carrier recovery, symbol demapping, and error-correction decoding. In receivers for MDM in MMF $(D > 2)$, most of these latter functions may be performed as in PDM receivers with straightforward modifications. However, compensation of linear channel effects, especially MD and modal crosstalk, will require different architectures than in PDM receivers to achieve similar adaptation speed and complexity per information bit. The changes from PDM to MDM receivers are necessitated by the increased MIMO dimensionality $(D \times D$ versus $2 \times 2)$ and because MD in MMF has a delay spread far larger than PMD in SMF $(N_{MD} >> N_{PMD})$.

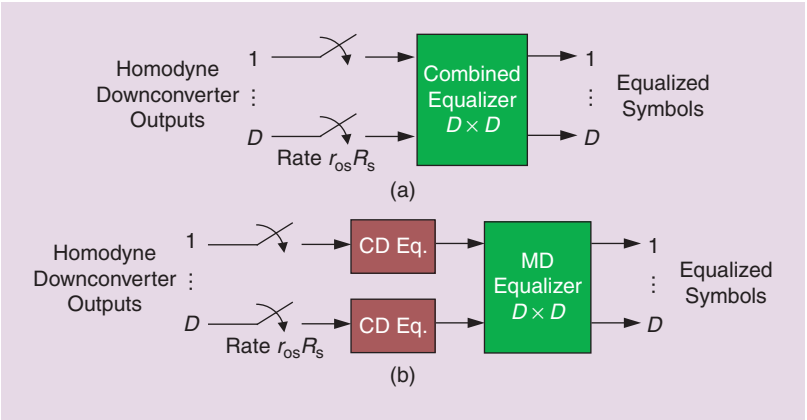
Several general considerations apply equally to PDM or MDM receivers:

- An equalizer for single-carrier modulation may use either FDE or time-domain equalization (TDE). FDE can handle long delay spreads (large N_{CD} or N_{MD}) with lower computational complexity than TDE by using the fast Fourier transform (FFT) for efficient implementation of convolution.
- CD is nearly fixed, changing very slightly with temperature variations, so it can be compensated by a programmable (but not adaptive) equalizer. PMD (or MD) can change as fast as the 10–100 μs time scale [10], [11] and so requires an adaptive $D \times D$ MIMO equalizer.
- Equalization of CD and MD (PMD in the special case $D = 2$) can be performed together or separately. Figure 6(a) shows a combined $D \times D$ equalizer for CD and MD (or PMD). Figure 6(b) shows a set of D equalizers for CD, followed by a $D \times D$ equalizer for MD (or PMD).

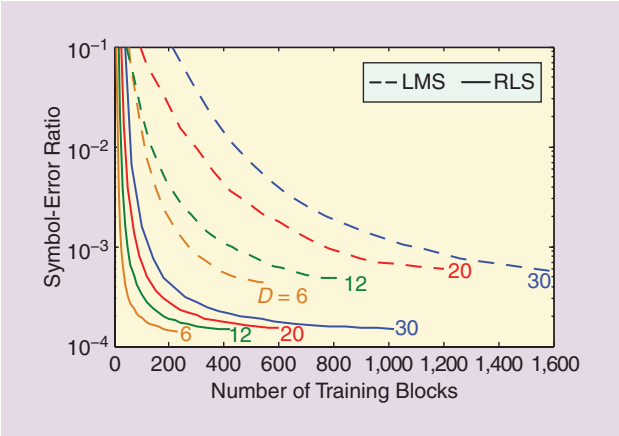
These considerations, along with differences in delay spread and MIMO dimensionality, provide the rationale for choosing different equalizer architectures for PDM and MDM receivers.

Most receivers for PDM in SMF $(D = 2)$ use the architecture of Figure 6(b) with a hybrid of FDE and TDE. Since N_{CD} is large, two programmable FDEs compensate CD efficiently. Since N_{PMD} is small, an adaptive 2×2 TDE compensates PMD and polarization crosstalk efficiently and facilitates fast adaptation [28].

In MDM systems, both N_{CD} and N_{MD} are large, so TDE is not viable for either CD or MD, but the optimal FDE architecture is not obvious. Combined equalization of CD and MD using an adaptive $D \times D$ FDE, as in Figure 6(a), has the lowest computational complexity for equalization (not considering adaptation) [19]. But



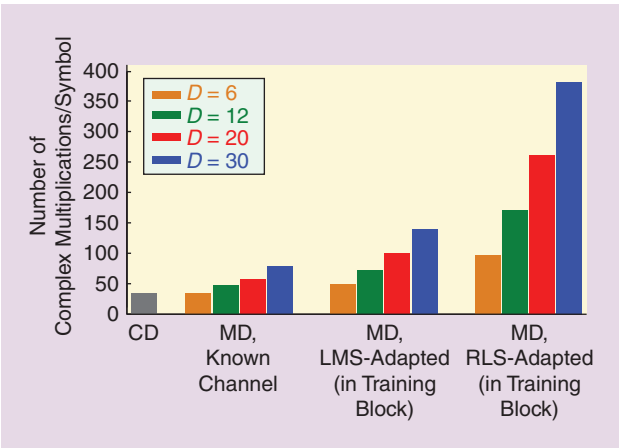
[FIG6] Single-carrier MIMO equalizer structures: (a) a combined equalizer for CD and MD (or PMD) and (b) separate equalizers.



[FIG7] The average symbol-error ratio versus the number of training blocks for $D = 6$, 12, 20, and 30 modes with RLS (solid line) and LMS (dashed line) algorithms (SNR = 10.5 dB, no MDL, $\kappa = 0.999$, and $\mu = 1.5 \cdot 10^{-5}$).

the FFT block length N_{FFT}, which should scale in proportion to $N_{\mathrm{CD}} + N_{\mathrm{MD}}$, becomes long, slowing down adaptation and impeding a single-chip realization. The separate equalizer structure of Figure 6(b) represents a more practical alternative. CD is compensated by a set of D programmable FDEs with FFT block length N_{FFT} proportional to N_{CD}, while MD and modal crosstalk (and any residual CD) are compensated by an adaptive $D \times D$ FDE with FFT block length N_{FFT} dependent on N_{MD} and D. While Figure 6(b) requires a slightly higher complexity than Figure 6(a) for equalization alone [19], minimizing the FFT block length in the adaptive $D \times D$ FDE facilitates fast adaptation and a single-chip realization.

There are two well-known approaches for FFT-based FDE for single-carrier modulation. Block convolution, e.g., overlap-save, avoids any overhead associated with a cyclic prefix, but complicates the realization of an adaptive FDE, requiring additional FFTs to enforce time-domain gradient constraints [29]. An alternate approach, adopted here, is to prepend a cyclic prefix of length N_{CP}



[FIG8] The number of complex multiplications/symbol for various frequency-domain equalization structures for $D = 6$, 12, 20, and 30 modes.

to each block of $N_{\mathrm{FFT}}/r_{\mathrm{os}}$ symbols before transmission. The cyclic prefix length N_{CP}, must be no shorter than the channel delay spread (N_{CD}, N_{MD} or $N_{\mathrm{CD}} + N_{\mathrm{MD}}$). When using a cyclic prefix, linear filtering by the channel corresponds to a multiplicative relationship between the FFTs of input and output signal blocks, which simplifies realization of FFT-based FDE. While cyclic prefix transmission decreases complexity, it reduces system throughput and average-power efficiency, as quantified by an efficiency parameter $\eta_{\mathrm{CP}} = N_{\mathrm{FFT}}/(N_{\mathrm{FFT}} + N_{\mathrm{CP}})$. Given a channel delay spread defining N_{CP}, η_{CP} is maximized by choosing $N_{\mathrm{FFT}} >> N_{\mathrm{CP}}$. Single-carrier modulation with FDE aided by a cyclic prefix or unique word has become popular in wireless systems [30].

ADAPTIVE EQUALIZATION

Assuming the architecture of Figure 6(b), we discuss adaptive frequency-domain implementation of the $D \times D$ equalizer for compensating MD, comparing the LMS and RLS algorithms. The linear MIMO channel $\mathbf{M}(\Omega)$ corresponds to a $D \times D$ matrix $\mathbf{M}[k]$ in the discrete frequency domain. The equalizer is described by a $D \times D$ matrix $\mathbf{W}[k]$. Since a typical channel matrix $\mathbf{M}[k]$ is close to unitary, either LMS or RLS can yield a $\mathbf{W}[k]$ close to the inverse of $\mathbf{M}[k]$. Both LMS and RLS are designed to iteratively minimize the squared error at all discrete frequencies. LMS is a stochastic gradient descent minimization using instantaneous error estimates [29], described by an update equation for the equalizer at frequency k

$$\mathbf{W}[k] \leftarrow \mathbf{W}[k] + \mu(\tilde{\mathbf{x}}[k] - \mathbf{W}[k]\tilde{\mathbf{y}}[k])\tilde{\mathbf{y}}[k]^H, \tag{9}$$

where $\tilde{\mathbf{x}}[k] = [X_1[k] \ \ldots \ X_D[k]]^T$ represents an FFT of sampled known training or detected data blocks, $\tilde{\mathbf{y}}[k] = [Y_1[k] \ \ldots \ Y_D[k]]^T$ represents an FFT of the sampled received signals, and μ is the step size. RLS involves iterative minimization of an exponentially weighted cost function [29], described by update equations

$$\mathbf{W}[k] \leftarrow \mathbf{W}[k] + (\tilde{\mathbf{x}}[k] - \mathbf{W}[k]\tilde{\mathbf{y}}[k])\tilde{\mathbf{y}}[k]^H(\mathbf{R}[k]\kappa^{-1}), \tag{10}$$

$$\mathbf{R}[k] \leftarrow \left(\mathbf{R}[k]\kappa^{-1}\right) - \frac{\left(\mathbf{R}[k]\kappa^{-1}\right)\tilde{\mathbf{y}}[k]\tilde{\mathbf{y}}[k]^H\left(\mathbf{R}[k]\kappa^{-1}\right)}{1 + \tilde{\mathbf{y}}[k]^H\left(\mathbf{R}[k]\kappa^{-1}\right)\tilde{\mathbf{y}}[k]}, \tag{11}$$

where $\mathbf{R}[k]$ is an tracked inverse time-averaged weighted correlation matrix [24] and κ is a forgetting factor.

The discrete frequencies for FDE are determined by the symbol rate R_s, oversampling rate r_{os}, and FFT block length N_{FFT}, while the cyclic prefix length must be at least equal to the MD memory length N_{MD} [24]. The choice of N_{FFT} has major consequences on the system, as it affects the adaptation time, cyclic prefix efficiency, and complexity. Given n_{tr}, the number of training blocks required until convergence, the adaptation time is $T_{\mathrm{adapt}} = n_{\mathrm{tr}}(N_{\mathrm{FFT}} + N_{\mathrm{MD}})T_s/r_{\mathrm{os}}$, so fast adaptation favors small N_{FFT}. On the other hand, a high cyclic prefix efficiency η_{CP} favors a large N_{FFT}. Overall computational complexity, including equalization and adaptation, is hardly affected by the choice of N_{FFT} over the range yielding acceptable T_{adapt} and η_{CP} [24].

To study adaptive FDE techniques, we consider the long-haul system described above, using 2,000 km (20 spans, each 100-km long) of GIGDC fiber, assume strong-mode coupling with section length $L_{\text{sec}} = 1$ km, neglect MDL, and assume symbol rate $R_s = 32$ Gbaud and oversampling rate $r_{\text{os}} = 2$. Initially, with the goal of achieving fast adaptation, we assume $N_{\text{FFT}} = 2^{11}, 2^{11}, 2^9, 2^9$ for $D =$ 6, 12, 20, and 30, respectively, which yields a fairly low cyclic prefix efficiency $\eta_{\text{CP}} \sim 80\%$. Later, we consider increasing N_{FFT} to increase η_{CP} toward 100%. We assume quaternary phase-shift keying and do not include error-correction coding. We quantify signal-to-noise ratio (SNR) by the transmitted signal power per mode divided by the noise variance per mode at the receiver, choosing SNR = 10.5 dB. We present Monte Carlo simulations averaged over random channel and symbol realizations.

Figure 7 illustrates equalizer convergence in terms of symbol-error ratio (SER) versus training block number. Although N_{MD} decreases with increasing D for these GIGDC fibers, the equalizer dimensionality scales as D^2, so the number of training blocks required for convergence increases with increasing D. 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 $\sim 6D$ and $\sim 15D$, respectively. For LMS, the knees occur at $\sim 20D$ to $\sim 30D$, while the flat-SER regime starts at $\sim 50D$ to $\sim 70D$. Choosing the number of training blocks n_{tr} as the onset of the flat-SER region, the estimated adaptation time T_{adapt} is ~ 3–$5\,\mu$s for RLS and ~ 15–$25\,\mu$s for LMS. These should be compared with 25 μs, the experimentally estimated time scale for mode-coupling dynamics [10]. While LMS appears to converge fast enough, it suffers from low cyclic prefix efficiency $\eta_{\text{CP}} \sim 80\%$ and higher asymptotic SERs. On the other hand, the convergence of RLS appears to be faster than required, so N_{FFT} can be quadrupled, yielding $\eta_{\text{CP}} \sim 95\%$. Moreover, RLS yields lower asymptotic SERs than LMS.

Another important advantage of RLS over LMS is its robustness for ill-conditioned channel matrices caused by MDL. In the presence of MDL, RLS converges reliably, albeit to a higher asymptotic SER than in the absence of MDL. By contrast, LMS must use a reduced step size to ensure convergence, which slows down convergence, and its asymptotic SER is degraded more than that of RLS [24]. Another advantage is that as the SNR increases, the asymptotic SER for RLS decreases more rapidly than that for LMS [24].

The major drawback of RLS is the high adaptation complexity per step arising from the matrix operations in (10) and (11). Figure 8 compares computational complexities, in terms of complex multiplications per data symbol in a training block, for FDEs adapted using LMS or RLS. This complexity scales sublinearly with D for both algorithms, and is 2.0 to 2.7 times higher for RLS than LMS for a given value of D. However, RLS converges to a low SER using ~ 4–5 times fewer training blocks, so the overall complexity for initial adaptation to an unknown channel is lower for RLS than for LMS. For reference, Figure 8 also shows complexities for equalizing CD (using overlap-save [19]) and MD for a known channel (using a cyclic prefix [24]), assuming optimized block lengths.

DISCUSSION

We have considered adaptation to an initially unknown channel, which is readily modeled without making assumptions about channel dynamics. Continuously tracking a channel in decision-directed mode is critically important. Our simulations indicate that continuous RLS adaptation would be effective, but at the cost of potentially prohibitive complexity and power consumption. Alternatives for tracking include RLS adaptation at periodic intervals or continuous LMS adaptation. Evaluating tracking performance requires models for channel dynamics, such as one proposed for fibers with two mode groups ($D = 6$) [31]. Such models should be extended to larger D and validated experimentally.

CONTINUOUS RLS ADAPTATION WOULD BE EFFECTIVE, BUT AT THE COST OF POTENTIALLY PROHIBITIVE COMPLEXITY AND POWER CONSUMPTION.

GD spread is a key factor determining the complexity, adaptation speed and cyclic prefix efficiency of MIMO equalizers for MDM systems. To manage GD spread, we have proposed fibers with low uncoupled GD spread in conjunction with strong mode coupling described by a section length $L_{\text{sec}} = 1$ km. As noted in [22], [26], manufacturing process variations may increase the uncoupled GD spread beyond its ideal value, while splices between fiber sections may be expected to lead to section lengths L_{sec} of order 5 km. 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 [32]. Fiber designs for reduced GD spread and methods to enhance mode coupling without increasing loss and MDL are important topics for future study.

Given the present uncertainty about mode-coupling dynamics and coupled GD spreads achievable in long-haul fibers, the MIMO FDE design parameters and performance and efficiency metrics provided here should be considered illustrative examples more than precise determinations.

In this article, we have focused on MIMO equalization, which is one of several important signal processing functions in MDM systems. Optimization of other functions, such as carrier recovery, and the integration and hardware implementation of the various functions, are important topics for future works.

ACKNOWLEDGMENTS

We are grateful to the guest editors for inviting us to write this tutorial article and to the anonymous reviewers for their helpful suggestions. This research was supported by National Science

Foundation grant number ECCS-1101905, Corning, Inc., and a Stanford Graduate Fellowship.

AUTHORS

Sercan Ö. Arık (soarik@stanford.edu) received the B.S. degree in 2011 from Bilkent University (Ankara, Turkey) and the M.S. degree in 2013 from Stanford University in electrical engineering. He is currently working toward his Ph.D. degree at Stanford University. He worked at EPFL Signal Processing Labs (Lausanne, Switzerland) in 2010, Google (Mountain View, California) in 2012, and Mitsubishi Electric Research Labs (Cambridge, Massachusetts) in 2013. His main current research interests include multimode optical communications; advanced modulation, coding, and digital signal processing techniques for fiber-optic systems; and optical signal processing. He is a Student Member of the IEEE.

Joseph M. Kahn (jmk@ee.stanford.edu) is a professor of electrical engineering at Stanford University. His research addresses communication and imaging through optical fibers, including modulation, detection, signal processing, and spatial multiplexing. He received A.B. and Ph.D. degrees in physics from the University of California, Berkeley, in 1981 and 1986. From 1987 to 1990, he was at AT&T Bell Laboratories, Crawford Hill Laboratory, in Holmdel, New Jersey. He was on the faculty of the Electrical Engineering Department at the University of California, Berkeley, from 1990 to 2003. In 2000, he cofounded StrataLight Communications, which was acquired in 2009 by Opnext, Inc. He received the National Science Foundation Presidential Young Investigator Award in 1991 and is a Fellow of the IEEE.

Keang-Po Ho (kpho@ieee.org) received his Ph.D. degree from the University of California, Berkeley, in 1995. He is currently with Silicon Image as a senior manager for the Baseband Algorithm Group. In Silicon Image (SiBEAM before acquisition), he invented and implemented wireless HD video area network using 60-GHz antenna array, 3.8-Gb/s OFDM modem, and beam search. He was the chief technology officer and cofounder of StrataLight Communications. He had also been with IBM, Bellcore, CUHK, and National Taiwan University. He has authored over 200 journal and conference articles, 15 issued patents, and one book. He is a Senior Member of the IEEE.

REFERENCES

[1] R. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Labs Tech. J.*, vol. 14, no. 4, pp. 3–9, 2010.

[2] E. Ip, P. Ji, E. Mateo, Y. 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.

[3] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightwave Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

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

[5] S. Randel, R. Ryf, C. Schmidt, M. A. Mestre, R.-J. Essiambre, and P. J. Winzer, "MIMO processing for space-division multiplexed transmission," in *Proc. Advanced Photonics Congr.*, 2012, paper SpW3B.4.

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

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

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

[9] 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, The Netherlands: Elsevier, pp. 491–568, 2013.

[10] X. Chen, J. He, A. Li, J. Ye, and W. Shieh, "Characterization of dynamic evolution of channel matrix in two-mode fibers." in *Proc. Optical Fiber Communication Conf. and Expo./Nat. Fiber Optic Engineers Conf. (OFC/NFOEC)*, Anaheim, CA, Mar. 2013, paper OM2C.3.

[11] P. Krummrich and K. Kotten, "Extremely fast (microsecond timescale) polarization changes in high speed long haul WDM transmission systems," in *Proc. Optical Fibers Communication Conf.*, 2004, pp. 1–3.

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

[13] W. Shieh, H. Bao, and Y. Tang, "Coherent optical OFDM: Theory and design," *Opt. Express*, vol. 16, no. 2, pp. 841–859, Jan. 2008.

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

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

[16] S. Fan and J. M. Kahn, "Principal modes in multimode waveguides," *Opt. Lett.*, vol. 30, no. 2, pp. 135–137, Jan. 2005.

[17] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single-mode fibers," *J. Lightwave Technol.*, vol. 9, no. 11, pp. 1439–1456, Nov. 1991.

[18] 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.

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

[20] 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.

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

[22] D. Peckham, Y. Sun, A. McCurdy, and R. Lingle, "Few-mode fiber technology for spatial multiplexing," in *Optical Fiber Telecommunications VI A*, I. P. Kaminow, T. Li, and A. E. Willner, Eds. Amsterdam, The Netherlands: Elsevier, pp. 283–319, 2013.

[23] R. Ryf, M. A. Mestre, A. H. Gnauck, S. Randel, C. Schmidt, R.-J. Essiambre, P. J. Winzer, R. Delbeu, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, D. W. Peckham. A. McCurdy, and R. Lingle, "Low-loss mode coupler for mode-multiplexed transmission in few-mode fiber," in *Proc. Optical Fiber Communication Conf.*, Los Angeles, CA, Mar. 2012, paper PDP5B.5.

[24] S. O. Arik, D. Askarov, and J. M. Kahn, "Adaptive frequency-domain equalization in mode-division multiplexing systems," submitted for publication.

[25] D. Đonlagic, "Opportunities to enhance multimode fiber links by application of overfilled launch," *J. Lightwave Technol.*, vol. 23, no. 11, pp. 3526–3540, Nov. 2005.

[26] L. G. Nielsen, Y. Sun, J. W. Nicholson, D. Jakobsen, K. G. Jespersen, R. Lingle, and B. Palsdottir, "Few mode transmission fiber with low DGD, low mode coupling, and low loss." *J. Lightwave Technol*, vol. 30, no. 23, pp. 3693–3698, Dec. 2012.

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

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

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

[30] 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.

[31] 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.

[32] 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.