Successive Interference Cancellation on Frequency-Selective Channels With Mode-Dependent Gain

Elaine S. Chou and Joseph M. Kahn, *Fellow, IEEE*

***Abstract*—In ultra-long-haul optical communication systems, polarization-dependent gain from single-mode semiconductor optical amplifiers or mode-dependent gain from multi-mode erbium-doped fiber amplifiers causes a channel's achievable information rate (AIR) using linear minimum-mean-square error (MMSE) detection to become substantially lower than the capacity of optimal maximum likelihood detection owing to loss of orthogonality between modes. However, such capacity loss can be mitigated using techniques that retain reasonable complexity, such as successive interference cancellation (SIC). In wireless systems, multicarrier transmission with many subcarriers transforms the frequency-selective channel into a set of uncoupled memoryless subchannels, and it is possible to obtain good performance using SIC on each subchannel independently. Multicarrier transmission in optical fibers, however, is susceptible to performance loss caused by Kerr nonlinearity. Therefore, we study methods for applying SIC on frequency-selective optical fiber channels using single-carrier transmission. We compare the theoretical performance of MMSE and SIC equalizers on frequency-selective optical fiber channels operating at 64 Gbaud with transmission distances up to 15,000 km and evaluate the penalties of SIC implementation. We consider single-mode fiber links with random artificial polarization-mode dispersion inserted along the link to enhance frequency diversity, as well as multi-core fiber links supporting 14 spatial and polarization modes. We show that SIC can increase outage capacity by 19%, but the AIR increase is reduced by almost half if error correction coding is not performed across modes. We confirm that SIC can adapt to fast channel perturbations with only a 3% decrease in AIR.**

***Index Terms*—Adaptive equalizers, frequency-selective channels, interference cancellation, MIMO, optical fiber communication.**

I. Introduction

AMONG currently deployed ultra-long-haul submarine links, single-mode fibers (SMFs) with erbium-doped fiber amplifiers (EDFAs) providing periodic optical amplification are prevalent. Nevertheless, advances in fiber and amplifier technologies provide viable alternatives to single-mode EDFA-amplified systems. In particular, semiconductor optical amplifiers (SOAs) have the potential to serve as low cost, wide bandwidth alternatives to EDFAs [1], [2] in single-mode systems, and multimode fibers (MMFs) or multi-core fibers (MCFs) paired with multi-mode EDFAs enable space-division multiplexing (SDM) for higher capacity density. While new designs continue to lower the mode-dependent gain (MDG)[1] in SMF and MDG in MMF or MCF. of SOAs and multi-mode EDFAs, they still exhibit higher MDG than single-mode EDFAs [3]–[6]. MDG has been shown to limit the performance of ultra-long-haul optical communication systems [7]–[9], reducing the capacity for ideal maximum likelihood (ML) detection. Additionally, on a channel with MDG, minimum-mean-square error (MMSE) detection results in lower instantaneous achievable information rate (AIR) and higher AIR variance than ML detection. The penalties to linear MMSE detection introduced by MDG motivate the study of equalization techniques offering better performance than linear MMSE detection. For example, successive interference cancellation (SIC) in conjunction with forward error correction (FEC) to eliminate errors between decoding layers can achieve performance approaching optimal ML detection [10].

In SIC, each layer of interference cancellation equalizes one data stream transmitted into a fiber mode using a multiple-input, single-output filter, performs FEC decoding, and subtracts the influence of the decoded data stream from the received signal. MDG in conjunction with mode coupling and dispersion results in a frequency-dependent channel response. In wireless systems, where SIC was first applied, SIC is extended to frequency-selective channels using multicarrier transmission to decompose the channel with memory into numerous memoryless subchannels to which SIC can be applied separately. In optical fibers, however, transmission using a large number of subcarriers is susceptible to performance loss caused by Kerr nonlinearity [11]. Hence, this paper focuses on SIC implementation compatible with single-carrier transmission, but our results can be extended to systems using a small number of carriers per transceiver, which can be practical as well [12].

Various forms of space-time coding and interference cancellation have been demonstrated on wireless and optical fiber channels with MDG [13]–[16], and recently,

Manuscript received July 10, 2021; revised November 20, 2021 and December 27, 2021; accepted January 27, 2022. Date of publication February 11, 2022; date of current version June 16, 2022. This work was supported by the Rambus Corporation Stanford Graduate Fellowship. *(Corresponding author: Elaine S. Chou.)*

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2022.3150355.

Digital Object Identifier 10.1109/JLT.2022.3150355

[1]Throughout this paper, the term "MDG" includes both polarization-dependent gain (PDG) in SMF and MDG in MMF or MCF.

a novel affine projection-based adaptation algorithm was demonstrated in conjunction with unreplicated SIC on a multiple-input multiple-output (MIMO) optical fiber channel with memory induced by MDG [17]. However, there has not been comprehensive study of the effect of broadband SIC equalizers on the AIR of links with distributed MDG, which have statistics different from the Rayleigh fading channel [10] encountered in wireless links and simulated in [16].

This paper studies the implementation of SIC on frequency-selective optical fiber links with MDG using single-carrier transmission. The performance gain of SIC over MMSE detection on long-haul optical links is quantified through average and outage capacity, evaluated over the ensemble of random instantiations of a channel with a specified standard deviation of MDG per amplifier. The remainder of this paper is structured as follows. Section II briefly reviews the multi-section channel model used in the study, discusses the potential for SIC to mitigate AIR loss in the presence of MDG, and defines the capacity metrics used to evaluate channel performance. Section III overviews single-carrier methods for implementing SIC on frequency-selective channels and addresses related implementation considerations such as complexity and implementation penalties. Section IV presents simulations of SIC on channels with MDG, expanding on previous results [18], and considers SIC on dynamic channels with MDG. Section V concludes the paper.

II. Background and Theoretical Framework

The transmission link studied includes fiber spans that induce coupling between modes, alternating with lumped amplifiers that induce MDG. At the receiver, equalization in the form of either MMSE detection or SIC is performed. The combination of the channel and receiver specification determines the AIR of the link.

A. Channel Model

A fiber is modeled as a concatenation of sections following the model introduced in [8] and further developed in [19]. Fiber section k has a transfer matrix that can be decomposed into unitary matrices $\mathbf{U}^{[k]}$ and $\mathbf{V}^{[k]}$ representing input and output mode coupling, and a diagonal matrix $\mathbf{\Lambda}^{[k]}$ that applies a different gain and group delay to each mode. Section k of a fiber with D total spatial and polarization modes is represented by a $D \times D$ transmission matrix [19]

$$\mathbf{H}^{[k]}(\omega) = \mathbf{V}^{[k]}\mathbf{\Lambda}^{[k]}(\omega)\mathbf{U}^{[k]H}, \tag{1}$$

where $(\cdot)^H$ denotes matrix Hermitian conjugate, and the diagonal matrix has the form

$$\mathbf{\Lambda}^{[k]}(\omega) = \mathrm{diag}\left[\exp\left(\mathbf{g}^{[k]}/2 - j\omega\boldsymbol{\tau}^{[k]}\right)\right], \tag{2}$$

where $\mathbf{g}^{[k]}$ and $\boldsymbol{\tau}^{[k]}$ are vectors of the gain in log power units and group delay in seconds, respectively, of each mode in section k; $\exp$ is elementwise exponentiation; and diag creates a diagonal matrix with the argument vector on the diagonal. Elements in $\mathbf{g}^{[k]}$ and $\boldsymbol{\tau}^{[k]}$ both sum to zero. The MDG of a channel is characterized by the standard deviation of the elements of $\mathbf{g}$, σ_g. The overall transfer matrix is constructed by multiplying the transfer matrices of all K sections and normalizing to unit average power, resulting in

$$\mathbf{H}(\omega) = \tfrac{1}{\alpha}\prod_{k=1}^{K}\mathbf{H}^{[k]}(\omega) = \tfrac{1}{\alpha}\mathbf{V}(\omega)\mathbf{\Lambda}(\omega)\mathbf{U}^{H}(\omega), \tag{3}$$

where the normalization factor α is constant across all channel realizations for a given standard deviation of MDG per amplifier σ_g [19]. For example, $\alpha = \exp(K\sigma_g^2/2)$ in an SMF link with K amplifiers each with standard deviation of PDG σ_g.

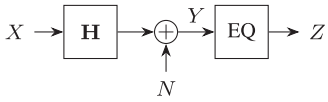

Fig. 1. Block diagram of discrete-time channel model. The multi-dimensional input X with is transmitted through the channel with transfer matrix $\mathbf{H}$, and white Gaussian noise N is added to produce the channel output Y. After the channel, equalization (EQ) produces the signal Z. The path from X to Z is assumed to have no memory and is characterized by the variance of interference and residual noise after equalization.

Transmit power is adjusted to optimize the optical signal-to-noise ratio (SNR), balancing the effect of amplified spontaneous emission from optical amplifiers and nonlinear interference from fiber propagation. As a result, the link SNR is inversely proportional to the number of fiber spans [19]–[21]. SNR is defined as the ratio of the total power in all modes divided by the noise power per mode. The transmitter is assumed to have no knowledge of the channel, so power is equally divided among modes. Noise including amplified spontaneous emission from amplification and nonlinear interference is approximated as additive white Gaussian noise (AWGN) [8].

The channel block diagram is shown by the path between X and Y in Fig 1. At each sampling time, the channel input and output are complex-valued D-dimensional vectors corresponding to the D spatial and polarization modes transmitted in the fiber. In the time domain, the elements of $X(t) \in \mathbb{C}^D$ are independent and identically distributed (i.i.d.) complex circularly Gaussian random variables with zero mean and variance equal to $1/D$ times the total transmit power P_s. Capacity analysis is most easily applied in the frequency domain, where the Fourier Transform is used to convert $X(t)$ to $X(\omega)$. To produce the channel output, at each discrete frequency, the channel response $\mathbf{H}(\omega) \in \mathcal{C}^{D\times D}$ is applied via matrix multiplication and noise $N(\omega) \in \mathcal{C}^D$ is added: $Y(\omega) = \mathbf{H}(\omega)X(\omega) + N(\omega)$. The frequency-selective channel response $\mathbf{H}(\omega)$ follows (1) – (3), and randomness in the channel is introduced in the random unitary coupling matrices $\mathbf{U}^{[k]}$ and $\mathbf{V}^{[k]}$, as well as in the Gaussian-distributed vectors of amplifier gain $\mathbf{g}^{[k]}$ and group delay $\boldsymbol{\tau}^{[k]}$. $N(\omega)$ is a vector of i.i.d. zero-mean complex circularly Gaussian random variables with noise power $\sigma_n = P_s/(\mathrm{SNR})$. When calculating capacity metrics, is it assumed that equalization of the received signal removes the effects of channel memory, and the resulting equalizer output is Gaussian-distributed with residual AWGN noise characterized by the MMSE $= \min_W \|X - W(Y)\|^2$, where W is an equalization function and $\|\cdot\|$ is the L2 norm.

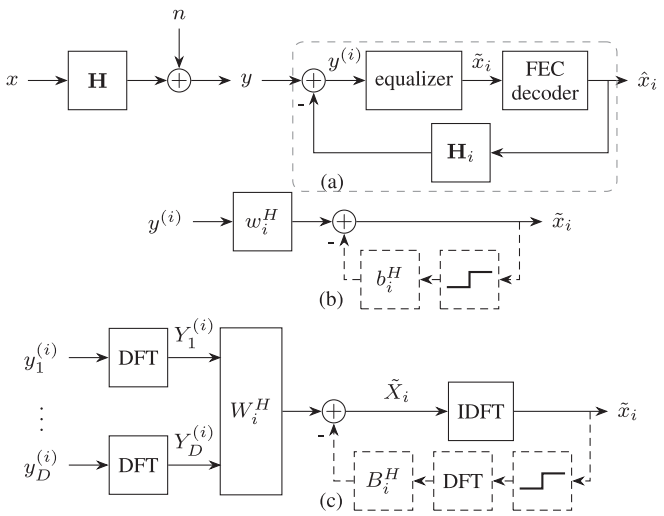

Fig. 2. Block diagrams describing SIC. (a) Entire transmission link from the transmitted signal x to the decoded symbols of the data stream transmitted on mode i, denoted by $\hat{x}_i$. The SIC block is enclosed in gray dashed lines. After the channel $\mathbf{H}$ with MDL, AWGN n is added. Equalization produces an estimate $\tilde{x}_i$. SIC removes the influence of transmitting the decoded data stream $\hat{x}_i$ through $\mathbf{H}_i$ to produce a modified received signal $y^{(i)}$ after each data stream is decoded. (b) Time-domain and (c) frequency-domain implementations of the equalizer block, showing optional decision-feedback equalizer (DFE) components in black dashed lines. In the time domain, feedforward filters w_i and feedback filters b_i equalize each data stream. In the frequency domain, each received data stream y_i is converted to the frequency-domain representation Y_i by a DFT before feedforward filtering by W_i to produce the estimate $\tilde{X}_i$. Thresholding is performed in the time domain, and the feedback filter B_i cancels postcursor intersymbol interference in the frequency domain.

B. Receiver Equalization and Achievable Information Rates

MMSE detection is widely used in many transmission systems. In the presence of MDG, extending MMSE detection to SIC takes advantage of the knowledge of already decoded data streams. SIC decodes the received signal iteratively, decoding one data stream in each layer. Interference from the decoded data stream is removed before filtering at the next layer, as shown in the feedback loop of Fig. 2(a) enclosed in the gray dashed box. Incrementing up from $i = 1$, the equalizer block filters $y^{(i)}$, the input to SIC layer i, to estimate the signal transmitted on mode i. For each vector of received symbols $y^{(i)}$, after the feedforward path equalizes and decodes the next symbol $\hat{x}_i$, the feedback path performs interference cancellation by subtracting the influence of the last decoded data stream to provide the input to the next SIC layer.

MMSE-SIC, which we employ in this study, uses MMSE filters to maximize the SNR at each layer. Let MMSE_i denote the MMSE of data stream i when data streams transmitted on all other modes are treated as noise. After the interference from the signals transmitted on modes 1 to $i-1$ has been removed, the MMSE reduces to $\text{MMSE}^{(i)}$. The first layer of MMSE-SIC is the same as MMSE detection of a single data stream. Using a Gaussian-distributed input in each element of the channel input X to maximize mutual information in the presence of Gaussian-distributed noise, the AIR of the first mode, treating interference from all other modes as noise and assuming that the equalized signal has no residual memory, is

$$\begin{aligned} C_{\text{sic}}^{(1)} &= -\log_2 \text{MMSE}^{(1)} \\ &= -\log_2 \text{MMSE}_1 = C_{\text{mmse}}^{(1)} \\ &= I(X_1; Y), \end{aligned} \tag{4}$$

where X is a D-dimensional random variable representing the entire signal transmitted on all input modes, X_i represents the data stream transmitted on input mode i, Y represents the signal received from the D output modes, and $I(X;Y)$ is the mutual information between variables X and Y. $\text{MMSE}_i \geq \text{MMSE}^{(i)}$, with equality when modes 1 to $i-1$ cause no interference to mode i. Because the already decoded data streams are cancelled, subsequent modes have AIRs given by

$$\begin{aligned} C_{\text{sic}}^{(i)} &= -\log_2 \text{MMSE}^{(i)} \\ &= I(X_i; Y | X_1, \ldots, X_{i-1}) \\ &\geq -\log_2 \text{MMSE}_i = C_{\text{mmse}}^{(i)} = I(X_i; Y). \end{aligned} \tag{5}$$

where analogously to MMSE detection,

$$\text{MMSE}^{(i)} = \left[\left(\mathbf{I} + \tfrac{\text{SNR}}{D} \mathbf{H}^{(i)H} \mathbf{H}^{(i)} \right)^{-1} \right]_{11}. \tag{6}$$

Here, $\mathbf{H}^{(i)} = [\mathbf{H}_i \cdots \mathbf{H}_D]$ is a submatrix of the channel transfer matrix $\mathbf{H}$ with columns $\mathbf{H}_1$ to $\mathbf{H}_{i-1}$ corresponding to cancelled input data streams removed, and $I(X;Y|Z)$ is the mutual information between X and Y conditioned on a third variable Z. The total AIR of a system with D modes is the sum of the AIRs of all modes.

$$\begin{aligned} C_{\text{sic}} &= \sum_{i=1}^{D} C_{\text{sic}}^{(i)} \\ &= \sum_{i=1}^{D} I(X_i; Y | X_1, \ldots, X_{i-1}) \\ &= I(X; Y) \\ &= C_{\text{mdg}}, \end{aligned} \tag{7}$$

where the third equality follows from the chain rule of mutual information and C_{mdg} is the capacity of a link with MDG using optimal ML detection. Therefore, on channels with MDG, MMSE-SIC can theoretically achieve performance approaching that of optimal ML detection. (7) does not depend on the order in which the summation is performed, so in theory, when all symbols are correctly decoded, the order in which data streams are decoded does not affect the achievable rate. In practice, however, iteratively decoding the remaining data stream with highest SNR relaxes the requirements on FEC performance by maximizing the SNR of the noisiest data stream [22].

To address the randomness in the transmission link, for each channel and equalization configuration, we evaluate two capacity statistics: average capacity and outage capacity. Average capacity is the expected value of AIR over all channel realizations. Given an outage probability q, the outage capacity is the information rate such that a fraction q of channel realizations have AIR below this value [23]. When evaluating outage capacity, we set the outage probability to $q = 10^{-3}$. We refer to the difference

between average and outage capacities as the outage capacity reduction.

C. Frequency Selectivity on Channels With MDG

In SDM systems, modal coupling and differential group delay cause an overall channel response that is dependent on frequency. The differential group delays between spatial modes in SDM systems are larger than the differential delay between the two polarization modes in SMF, so the effect of the frequency-dependent phase shift applied in (2) is larger as well. On its own, modal dispersion has a unit-gain frequency response, but coupling and dispersion modulate the amplitude of the frequency response, creating a frequency-selective channel. To generalize the AIR equations to a frequency-selective channel, the transmission matrix $\mathbf{H}$ and corresponding MMSEs are modeled as frequency dependent, and (5) is averaged in frequency across the bandwidth of the signal. In a system with strong frequency dependence, the channel's decorrelation bandwidth is small, so many independent channel realizations are included in the frequency average. Averaging over independent AIR values pulls the instantaneous AIR closer to the average capacity. In SMF links, which have low inherent frequency dependence, the outage capacity reduction is significant, increasing substantially if an MMSE receiver is used in place of an optimal receiver. The frequency dependence of SMF links can be increased by intentionally distributing random extra polarization-mode dispersion (PMD) along the link to induce frequency diversity [24].

To better understand the effect that frequency selectivity has on capacity of a channel with MDG, we plot the distribution of MMSE AIR versus optimal capacity in an SMF link across different transmission distances and artificial PMD values in Fig. 3. Random PMD is inserted into the link by adding differential group delay (DGD) between each fiber span, where the delays are drawn from a uniform distribution between $-\tau$ and τ. As expected, the MMSE AIR never exceeds the optimal capacity. For a short link, the capacity variance for optimal detection is very small, with a peak capacity difference less than 0.02 b/s/Hz across 10,000 channel realizations. In contrast, without the addition of artificial PMD, the MMSE AIR varies over a range of almost 2 b/s/Hz. As the transmission distance increases, the AIR spread increases for both MMSE and optimal detection. As seen from the blue points in Fig. 3, the optimal channel capacity has a clear minimum when no extra PMD is added. This minimum capacity corresponds to the channel realization with equal gain across all modes and frequencies and is given by [19]

$$C^{\min}_{\text{mdg}}(D, \text{SNR}) = D \log_2 \left(1 + \frac{\text{SNR}}{D\alpha^2}\right). \tag{8}$$

At this minimum capacity, modes remain orthogonal, so the MMSE AIR is equal to the optimal channel capacity. As the optimal channel capacity increases from the minimum, the MMSE AIR begins to spread out. If the modes remain orthogonal, there is no penalty for using MMSE detection, but as the interference between modes increases, so does the penalty for MMSE detection. The MMSE AIR for an SMF link is lower bounded

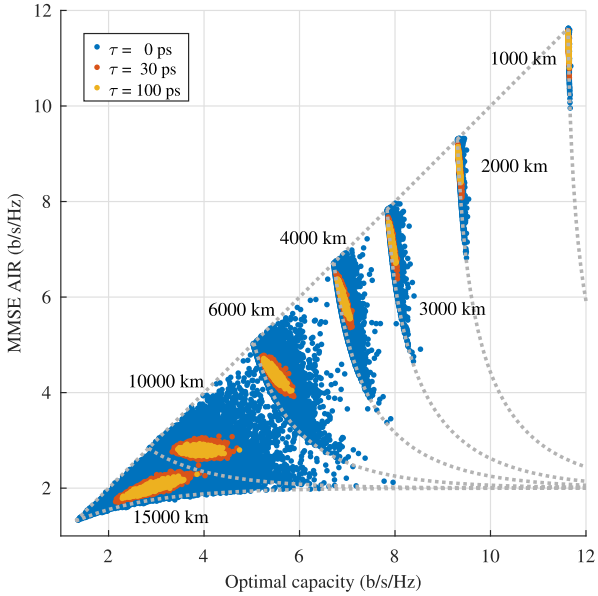

Fig. 3. Evolution of AIR for MMSE detection vs. capacity for optimal detection in SMF ($D = 2$ modes) as the link length grows from 1,000 km to 15,000 km. The standard deviation of PDG is $\sigma_g = 0.5$ dB per amplifier. For each link length, extra per-span PMD with DGD uniformly distributed between $-\tau$ and $+\tau$ is added between each 50 km span, with $\tau = 0$ ps (blue), 30 ps (red), and 100 ps (yellow). Dashed lines indicate theoretical bounds; lower bounds on MMSE AIR are given by (9).

by

$$C^{\text{lb}}_{\text{mmse}}(2, \lambda, \text{SNR}) = 2 - 2\log_2 \left[\left(1 + \frac{\text{SNR}}{2\alpha^2}\lambda\right)^{-1} + \left(1 + \frac{\text{SNR}}{2\alpha^2}\frac{1}{\lambda}\right)^{-1}\right], \tag{9}$$

where $\boldsymbol{\lambda}/\alpha^2$ and $1/(\boldsymbol{\lambda}\alpha^2)$ are the eigenvalues of $\mathbf{H}\mathbf{H}^H$. This lower bound is equal to $C^{\min}_{\text{mdg}}(2, \text{SNR})$ when $\lambda = 1$. In the limit of large mode gain imbalance, the lower bound approaches

$$\lim_{\lambda \to 0} C^{\text{lb}}_{\text{mmse}}(2, \lambda, \text{SNR}) = 2. \tag{10}$$

Even at high SNR, there is a nonzero, though admittedly very small, probability that the MMSE AIR will be as low as 2 b/s/Hz.

Without artificial PMD, the SMF frequency response is relatively flat over frequency, but distributing additional PMD along the link introduces frequency diversity that effectively averages capacity across frequency. The effect of frequency diversity can be observed in the concentration of points in the scatter plot around the mean of the distribution with decreased capacity variance. As the extra DGD inserted after each fiber span increases, the variance of capacity decreases. There is a limit, however, to how much the capacity variance can be reduced, as observed in [24]. Beyond $\tau = 200$ ps of per-span PMD with DGD uniformly distributed between $-\tau$ and $+\tau$, the additional outage capacity increase is minimal. The outage capacity reduction is inversely proportional to the square root of the frequency diversity order [25], so these reduced marginal gains from increasing PMD are expected. In SDM systems, in addition to frequency diversity, modal diversity [19], [26] contributes to averaging capacity across spatial channels. For a channel with $D = 14$ spatial and polarization modes, modal

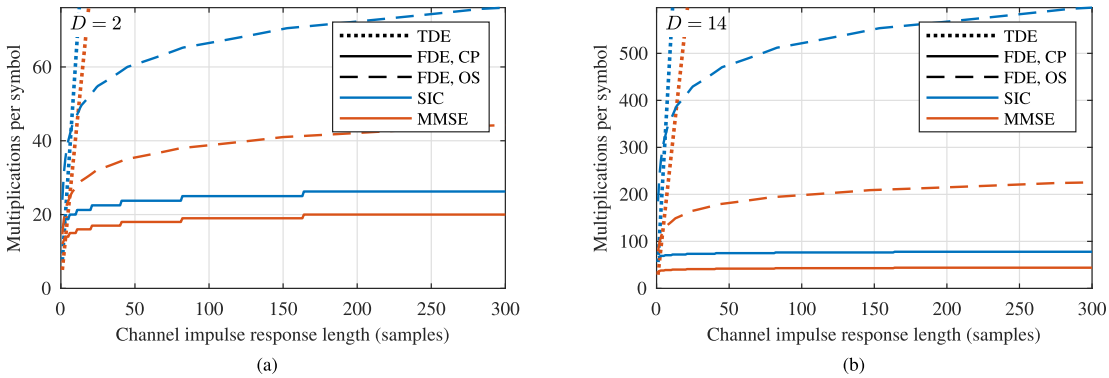

Fig. 4. Number of complex multiplications per symbol vs. channel impulse response length for TDE (dotted), FDE with CP (solid), and FDE with OS convolution (dashed). The length of the equalizer filters w_i is set equal to the impulse response length of the channel for TDE. The CP has 1% transmission overhead. The OS convolution overlap length is optimized to minimize complexity. Complexities for (a) $D = 2$ and (b) $D = 14$ spatial and polarization modes are shown. The FFT block length is forced to a power of 2.

diversity is sufficiently high that even at a single frequency, the distributions of MMSE AIR and optimal capacity exhibit the effects of averaging.

III. Implementation Considerations

The equalizer block in Fig. 2(a) can be implemented as shown in Fig. 2(b) using a decision-feedback equalizer (DFE) [13] in the time domain, or as shown in Fig. 2(c), in the frequency domain. SDM submarine links thousands of kilometers in length can have memory lengths on the scale of nanoseconds [27], corresponding to equalizer lengths of hundreds of symbols. In such channels requiring many equalizer taps, frequency-domain equalization (FDE) using the discrete Fourier transform (DFT) [28] is advantageous over time-domain equalization (TDE). FDE requires DFT and inverse DFT (IDFT) pairs to convert blocks of symbols between the time and frequency domains. But in exchange, filtering by convolution is replaced by scalar multiplication at each frequency. We consider two methods to implement filtering in the frequency domain: the cyclic prefix (CP) method [10], and the overlap-save (OS) method [29].

A. Complexity

The AIR increase offered by SIC over MMSE detection comes at a cost of increased computational complexity. SIC requires additional steps of calculating the effect of the channel on recovered symbols to subtract its influence. In the following analysis we focus on the complexity of SIC implemented using linear equalizers with filter weights adapted using the least-mean squares (LMS) algorithm and compare the complexity of SIC and MMSE, measured in number of complex multiplications. The equalization implementations under consideration are TDE, FDE using a CP, and FDE using OS convolution.

Fig. 4 compares the complexities of SIC and MMSE for the SMF and 7-core CC-MCF links, assuming that the DFT is computed using the radix-2 fast Fourier transform (FFT). The complexity of TDE increases rapidly with the channel impulse response length, and once the channel impulse response length reaches 10 symbols, the higher complexity of convolution over multiplication overtakes that of the DFTs needed for FDE. Although the CP method has lower complexity than OS convolution, it requires longer block lengths if small overhead is desired for a higher overall bit rate. Excessively long block lengths will impede tracking on a dynamic channel.

As an example, consider a 7-core coupled-core multi-core fiber (CC-MCF) link transmitting at $R_s = 64\,\mathrm{Gbaud}$ with transmission distance 10000 km. The group delay standard deviation of the fiber is $\sigma_\tau = 3.1\ \mathrm{ps}/\sqrt{\mathrm{km}}$, so the total standard deviation of group delay in the link is $\sigma = 310\ \mathrm{ps}$. The corresponding number of taps required with symbol-spaced sampling is about $N_{taps} = \lceil 5\sigma R_s \rceil = 100\ \mathrm{taps}$ [30]. In this case, FDE with a CP is the most computationally efficient option. OS convolution has complexity almost 7 times that of the CP method because of the number of DFTs required, which scales linearly with D. SIC has 1.8 times the complexity of MMSE detection when a CP is used. SIC implemented using FDE with OS convolution has 2.6 times the complexity of MMSE detection.

B. Implementation Penalties

FEC is needed between SIC layers to uphold the assumption of error-free symbol decisions used in AIR calculations, and it introduces delay after each SIC layer, resulting in latency that is linear in the number of modes D. Without FEC between SIC layers, error propagation in links with low SNR can cause SIC to perform more poorly than MMSE unless the SIC algorithm is adjusted to account for unreliable symbol decisions [31]. As the net coding gain of FEC codes used in optical communication systems increases, the resulting increase in tolerable pre-FEC bit-error ratio may cause error propagation in a decision-directed DFE. Error propagation can be avoided by using a linear equalizer, but not without incurring a cost in AIR. A linear equalizer will produce an MMSE equal to the average MMSE over the signal bandwidth B, corresponding to a decreased AIR of

$$C_{\mathrm{sic,le}} = -\log_2\left[\frac{1}{N}\sum_{k=0}^{N-1}\mathrm{MMSE}^{(i)}\left(\tfrac{2\pi(2k+1)B}{2N}\right)\right] \le -\frac{1}{N}\sum_{k=0}^{N-1}\log_2\left[\mathrm{MMSE}^{(i)}\left(\tfrac{2\pi(2k+1)B}{2N}\right)\right] = C_{\mathrm{mdg}}, \tag{11}$$

where the inequality holds by Jensen's inequality because $\log$ has negative concavity. $\mathrm{MMSE}^{(i)}(\omega)$, the MMSE of data stream i for a narrowband channel centered at frequency ω, is calculated using (6), evaluating $\mathbf{H}(\omega)$ at the center frequency of the frequency bin. An analogous expression for $C_{\mathrm{mmse,le}}$, the AIR of a channel equalized with linear equalizers and MMSE detection, is formed by replacing $\mathrm{MMSE}^{(i)}$ with MMSE_i in the right side of the first equality in (11). We designate the difference between the optimal outage capacity and achievable linear equalizer outage capacity as the linear equalizer (LE) penalty.

The AIRs considered thus far are achieved when the transmitted symbols follow a continuous Gaussian distribution. In practice, discrete modulation formats such as quadrature-amplitude modulation (QAM) are used, causing the achievable bit rate to decrease. To determine the effect of a discrete constellation order on AIR, we also consider the constrained capacity using uniform-probability QAM signals and bit-interleaved coding and decoding by evaluating the generalized mutual information (GMI) [32] for different QAM constellation orders. In contrast to capacity, GMI computes the mutual information between the transmitted bits and received symbols. Similarly to mutual information, the GMI for one channel instantiation can be computed from the MMSE values of the individual data streams. For each mode, GMI is computed based on transmission of signal with symbols drawn from a uniformly distributed, finite-sized discrete symbol alphabet over a single-mode AWGN channel with noise power corresponding to the MMSE of that mode, and GMI is summed over all modes.

Although vertical Bell Laboratories Layered Space-Time (V-BLAST) [22] is the most straightforward implementation of SIC, it is unable to achieve optimum outage capacity because transmit data streams are coded separately. Without coding across modes the AIRs given in (7) and (11) cannot be achieved, because the channel will be in outage if any of the D modes attempt to transmit at an information rate above that mode's AIR [10]. As a result, the outage capacity is limited by the minimum AIR across all modes, $\min_i C_{\mathrm{sic}}^{(i)}$. The outage capacity with outage probability q under separate coding of transmitted data streams decreases to the AIR such that a fraction q of channel realizations have at least one mode with AIR less than $1/D$ of the outage capacity. Other space-time codes that code across spatial dimensions, such as diagonal BLAST (D-BLAST) [33], are able to achieve optimum outage capacity. In exchange, D-BLAST suffers from transmission overhead because of the staggered transmission of codewords during initialization.

IV. Simulation Results

To determine average and outage capacities, we start by generating a set of 10,000 random instantiations of a channel with a specified MDG standard deviation σ_g at each amplifier, following the channel structure given in (1) – (3). We consider two fiber structures: SMF with $D = 2$ modes and 7-core CC-MCF with $D = 14$ modes. The intrinsic PMD standard deviation of the SMF and modal dispersion standard deviation of the CC-MCF are set to 0.05 ps/$\sqrt{\mathrm{km}}$ and 3.1 ps/$\sqrt{\mathrm{km}}$, respectively. The channel has MDG-inducing amplifiers every 50 km and carries a 64 Gbaud single-carrier signal. To determine the SNR at each transmission distance, we use the Gaussian-noise (GN) model of nonlinear propagation to model the effect of nonlinear interference as AWGN, optimizing launch power to maximize SNR in the presence of amplified spontaneous emission (ASE) noise and incoherent accumulation of nonlinear noise [19], [20]. Under the GN model, SNR is inversely proportional to the number of fiber spans. The SNR scaling is set such that a 2500 km link satisfies $\mathrm{SNR}/D = 14$ dB, following Q^2 measurements of quadrature phase shift keying (QPSK) transmission [27]. For an amplifier spacing of 50 km, the SNR at the end of a link of K spans is $\mathrm{SNR}(K) = \mathrm{SNR}(1)/K$, where $\mathrm{SNR}(1)/D = 31$ dB. Because of modal coupling during fiber propagation, even though the MDG standard deviations of all amplifiers are the same, the overall MDG of the link and consequently the capacity is different for each random instantiation [19]. For each channel instance and link length, the mutual information or GMI for each equalization method is computed based on the MMSE of each mode after equalization, as described in Sections II-B and III-B. For the purposes of computing average and outage capacities, the MMSE values used in the AIR equations are computed analytically based on the channel matrix, for example using (6) in the case of SIC. On a subset of channel instances, a discrete-time, random symbol sequence is generated and transmitted through the noisy channel, and adaptive equalization is performed to confirm agreement in MMSE values. FEC encoding and decoding are abstracted out, and in between SIC layers, we assume that FEC is able to correct all errors, so that only correct symbols are used in interference cancellation. Finally, the average and outage capacities are computed from AIRs of all channel instances. We compare the average and outage capacities for SIC and MMSE detection methods, considering DFE and linear equalizer implementations.

A. 7-Core Coupled-Core Multi-Core Fiber

First we study the 7-core CC-MCF ($D = 14$), investigating optimal SIC capacity, the effect of linear equalizers, contrained capacity with discrete signal constellations, and the effect of coding each data stream separately. These implementation penalties were analyzed in Section III-B. Fig. 5 shows the AIR comparison between MMSE detection and SIC for a 7-core CC-MCF link when the standard deviation of MDG is $\sigma_g = 0.3$ dB per amplifier. Consistent with the results of [19], all AIRs are roughly equal for a given equalization type (MMSE or SIC). For both SIC and MMSE detection, the outage capacity reduction and the LE penalty are small, both less than 0.05 b/s/Hz/spatial mode after 10000 km of fiber. The small LE penalty indicates that modal diversity alone is enough to keep the AIR close to that of ideal ML detection, consistent with the observations of Section II-C. There is a penalty for MMSE detection compared to SIC that increases with link length. A 10000 km link using SIC can achieve an AIR 0.7 b/s/Hz/spatial mode higher than the same link using MMSE detection, corresponding to a 19% increase in AIR.

Fig. 6 plots the constrained capacities of several uniform-probability QAM constellations for the same 7-core CC-MCF links as in Fig. 5. It also marks the reach of each QAM constellation using an ideal capacity-achieving FEC code with overhead restricted to 20%, the maximum recommended by OIF [34].

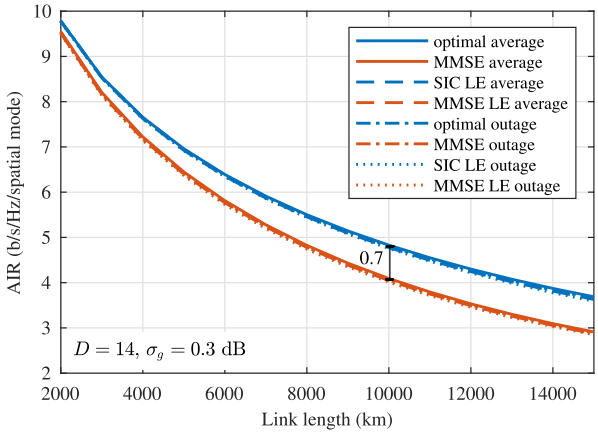

Fig. 5. Average and outage capacities vs. link length for a 7-core CC-MCF link with standard deviation of MDG $\sigma_g = 0.3$ dB per amplifier and 50 km fiber spans. Methods compared are SIC (blue) and MMSE (red) detection. AIRs using linear equalizers (LEs) with single-carrier transmission are similar to AIRs using multicarrier transmission or DFEs. Lines of the same color are overlapping, with less than 0.05 b/s/Hz/spatial mode difference.

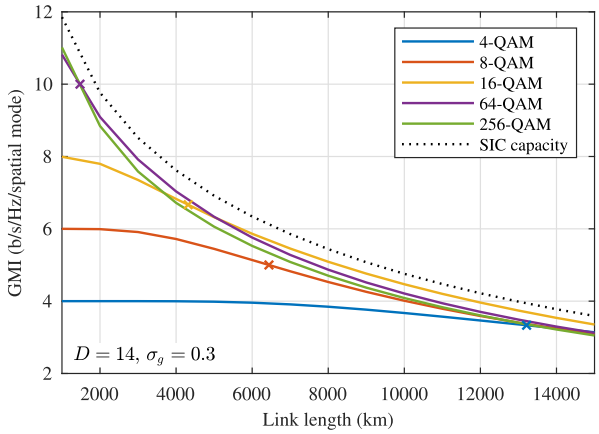

Fig. 6. Constrained outage capacity, quantified by GMI, vs. link length using SIC in a 7-core CC-MCF link. The standard deviation of MDG is $\sigma_g = 0.3$ dB per amplifier. Amplifiers are separated by 50 km fiber spans. Uniform QAM orders range from 4 to 256. Dotted line indicates the optimal unconstrained capacity. Crosses indicate the reach achievable by each QAM constellation using an ideal capacity-achieving FEC code with 20% overhead.

For 4-QAM, the reach is slightly above 13000 km. If higher FEC overhead is allowed, 16-QAM has the highest constrained capacity in links longer than 5000 km. There is a penalty of at least 1.6 b/s/Hz/spatial mode between the optimal capacity and the constrained capacity, due to the gap to capacity of the GMI for unshaped constellations. One option to decrease the gap is to use probabilistic shaping [35]. The decreased rate resulting from probabilistic shaping will also increase the transmission reach for a fixed maximum FEC overhead.

As explained in Section III-B, SIC architectures such as V-BLAST that code each data stream separately do not take full advantage of the modal diversity offered by an SDM channel. Fig. 7 shows the extent of capacity loss in 7-core CC-MCF using SIC when coding is performed separately on data streams transmitted on each mode, as opposed to across modes, as is assumed in the other AIR calculations. Without the transmit diversity gained from coding across different modes, the amount of modal diversity is insufficient to fully decrease the capacity

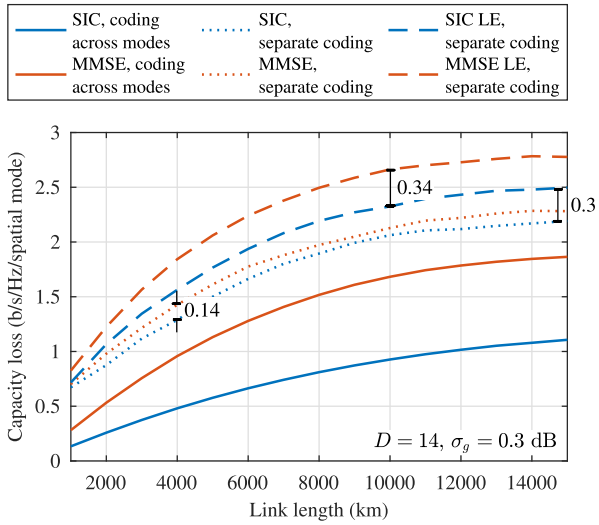

Fig. 7. Decrease in outage capacity associated with coding data streams separately in a 7-core CC-MCF with standard deviation of MDG $\sigma_g = 0.3$ dB per amplifier and 50 km fiber spans. Capacity loss is referenced to a link of equal SNR but no MDG. Methods compared are coding across modes (solid), separate coding of data streams transmitted on each mode (dotted), and separate coding of data streams with the linear equalizer (LE) penalty (dashed). AIRs for SIC and MMSE detection are shown in blue and red, respectively.

variance, so the outage capacity decreases. Additionally, there is an observable LE penalty that grows to 0.3 b/s/Hz/spatial mode for SIC at 15000 km. Notably, the outage capacity of MMSE becomes much closer to that of SIC, differing by less than 0.15 b/s/Hz/spatial mode along the entire link. The LE penalty of MMSE detection is slightly higher than that of SIC. If linear equalizers are used when data streams are coded separately, SIC increases outage capacity by 0.34 b/s/Hz/spatial mode compared to MMSE detection after 10000 km of fiber, an 11% increase. When data streams are coded separately, the channels with low capacity tend to have roughly equal performance across modes when using MMSE detection. Thus, although SIC decreases the MSE of data streams decoded later, the outage capacity is limited by the first, noisiest data stream. While the increased signal-to-interference-plus-noise ratio of data streams equalized with SIC does not always increase AIR, it still lowers the MSE, which offers practical advantages such as better adaptive filter convergence and decreased error propagation. As MDG increases, the capacity loss resulting from separate coding of data streams increases. After 10000 km of transmission, for a link with standard deviation of MDG $\sigma_g = 0.2$ dB per amplifier, the capacity loss is 1 b/s/Hz/spatial mode, compared to 1.4 b/s/Hz/spatial mode when $\sigma_g = 0.3$ dB and 1.9 b/s/Hz/spatial mode when $\sigma_g = 0.5$ dB per amplifier.

B. Single-Mode Fiber

For SMF ($D = 2$) links, which have inherently low frequency diversity, we focus on the effect of adding random artificial PMD throughout the link. Fig. 8 shows the average and outage capacities as a function of transmission distance for an SMF link with standard deviation of PDG $\sigma_g = 0.3$ dB per amplifier. The frequency response is relatively flat and the channel impulse response is only a few symbols in duration, leading to a small

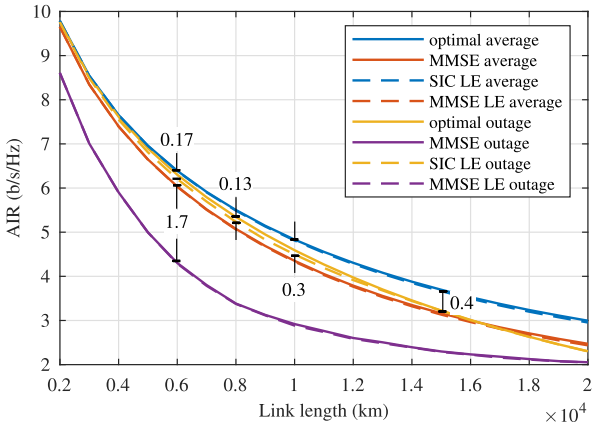

Fig. 8. Average and outage capacities vs. link length for SMF link with standard deviation of PDG $\sigma_g = 0.3$ dB per amplifier and 50 km fiber spans, comparing SIC (blue) and MMSE detection (red). Optimal capacities and MMSE AIRs can be achieved using multi-carrier transmission or DFEs in conjunction with SIC and MMSE, respectively. In contrast, linear equalizer (LE) AIRs are those achievable using single-carrier signals filtered only by LEs.

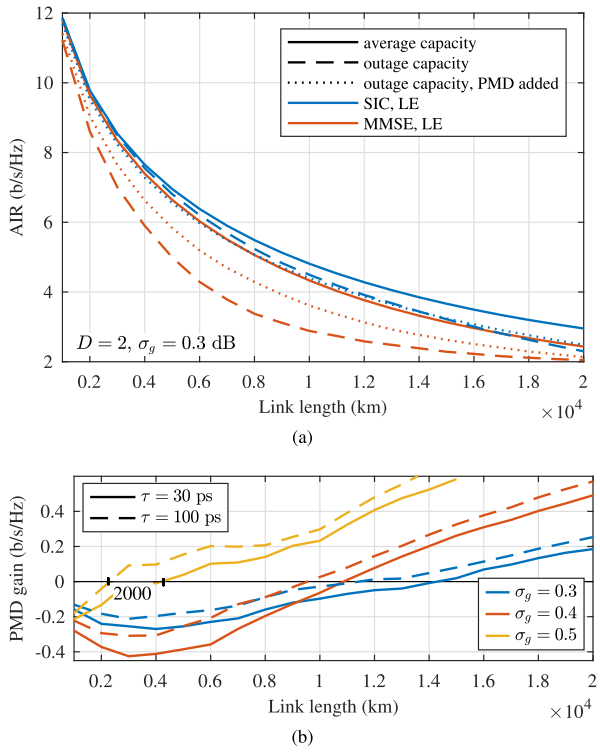

Fig. 9. Effect of artificial PMD in SMF links. (a) Average and outage capacities using SIC vs. link length for standard deviation of PDG $\sigma_g = 0.3$ dB per amplifier and 50 km fiber spans. Outage capacity is compared with and without extra PMD. Extra per-span PMD has DGD uniformly distributed between -30 and +30 ps. Methods compared include SIC (blue) and MMSE detection (red) using linear equalizers (LEs). (b) Outage capacity gained by SIC by adding extra PMD to a link with standard deviation of PDG $\sigma_g = 0.3$ dB (blue), 0.4 dB (red), 0.5 dB (yellow) per amplifier. After each 50 km fiber span, PMD with DGD uniformly distributed between $-\tau$ and $+\tau$ for $\tau = 30$ (solid) or $\tau = 100$ (dashed) ps is added.

LE penalty. The penalty for using MMSE detection is small if average capacities are compared but is much larger if outage capacities are considered instead. The LE penalty is slightly higher for SIC than MMSE detection, with a maximum penalty of 0.13 b/s/Hz or 2.5% at the intermediate distance of 8000 km. Even with a higher LE penalty, SIC can still increase the outage capacity significantly compared to MMSE detection, decreasing the outage capacity reduction by up to 1.5 b/s/Hz at 6000 km. However, a residual outage capacity reduction that cannot be eliminated using SIC alone remains. The SIC outage capacity reduction grows with link length even as the absolute AIR decreases. The outage capacity reduction is 0.3 b/s/Hz or 6% at 10000 km, increasing to 0.4 b/s/Hz or 12% at 15000 km.

While introducing artificial PMD universally increases outage capacity if optimal detection is used [19], this is no longer the case when linear equalizers are used. Fig. 9(a) plots average and outage capacities with SIC or MMSE implemented using linear equalizers as a function of transmission distance for an SMF link with standard deviation of PDG $\sigma_g = 0.3$ dB per amplifier. Fig. 9(b) plots the outage capacity increase of SMF links with SIC implemented using linear equalizers. If the effect of PDG is small, for example in shorter links or in links with low-PDG amplifiers, the penalty from increased channel memory length with artificial PMD outweighs its benefits in combating PDG. Although introducing artificial PMD decreases the capacity variance as desired, the LE penalty decreases average AIR. Consequently, at shorter distances, outage capacity can decrease compared to a link without artificial PMD.

For a link with standard deviation of PDG $\sigma_g = 0.3$ dB per amplifier and extra per-span PMD with DGD uniformly distributed between -30 and +30 ps, the crossover distance at which outage capacity with extra PMD becomes higher than that without PMD is 14000 km. If the standard deviation of PDG σ_g increases to 0.5 dB per amplifier, the crossover distance is below 5000 km. The crossover distance can also be decreased by increasing the amount of PMD introduced along the link, at the cost of longer memory length. For extra per-span PMD with DGD uniformly distributed between $-\tau$ and $+\tau$, increasing τ from 30 ps to 100 ps decreases the crossover distance by about 2000 km.

Although a channel with high PDG or long transmission distance is required for the introduction of artificial PMD to increase outage capacity, the benefit of artificial PMD in SMF links is more clear when considering the GMI of of discrete signal constellations with smaller constellation orders. Fig. 10 plots the constrained capacities of several uniform-probability QAM constellations for the same SMF links as in Fig. 9(a). When using QAM constellations and ideal FEC with 20% overhead in an SMF link with a standard deviation of PDG $\sigma_g = 0.3$ dB per amplifier, adding PMD increases the system reach for all QAM orders simulated. Higher gains in system reach are achieved for lower constellation orders, which are less sensitive to interference. Adding per-span PMD with DGD uniformly distributed between -30 and +30 ps increases the reach by 30% from 3000 km to 3900 km when transmitting a uniform 16-QAM constellation and more than 85% from 6300 km to 11700 km when transmitting 4-QAM. For link lengths within the system reach, extra PMD increases GMI.

C. Dynamic Channels and Adaptation

Because of environmental fluctuations, the transfer matrix of an optical fiber link changes over time, and an equalizer must

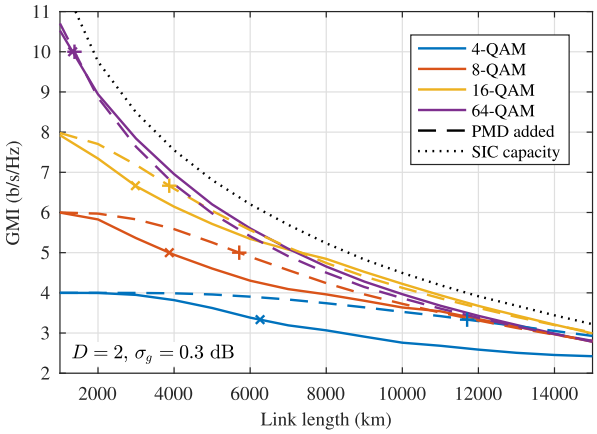

Fig. 10. Constrained outage capacity using SIC with linear equalizers, quantified by GMI, vs. link length in SMF. The standard deviation of PDG is $\sigma_g = 0.3$ dB per amplifier. Amplifiers are separated by 50 km of fiber. AIRs are compared with (dashed) and without (solid) extra PMD. Extra per-span PMD adds DGD uniformly distributed between -30 and +30 ps. Uniform QAM orders range from 4 to 64. Dotted line indicates the optimal unconstrained capacity. Crosses indicate the reach achievable by each QAM constellation using an ideal capacity-achieving FEC code with 20% overhead.

TABLE I
DYNAMIC CHANNEL PENALTIES

p	$D = 2$		$D = 14$	
	Stokes rotation rate (krad/s)	$C_{\text{sic,le}}$ loss	Stokes rotation rate (krad/s)	$C_{\text{sic,le}}$ loss
1×10^5	4	0.9%	30	1.2%
2×10^6	18	1%	100	12%
1×10^8	100	3%	400	–

Loss in $C_{\text{sic,le}}$ when adapting to a dynamic channel.
The parameter p controls perturbation strength in the dynamic channel model.
The symbol – indicates that decision-directed LMS could not converge on the dynamic channel.

successfully track channel dynamics without introducing too much excess MSE. In one submarine SMF link field measurement, the maximum Stokes vector rotation rate measured was 350 rad/s [36]. Fast channel dynamics can be modeled using a linear ramp model that adds a linear perturbation to the mode coupling matrices $\mathbf{V}^{[k]}$ and $\mathbf{U}^{[k]}$ of the multi-section model [37]. Two simplifying assumptions are made: that uncoupled modal gains and delays are constant over the perturbation period and that random mode coupling can change with time but is independent of frequency. The speed of the resulting rotation of the generalized Stokes vector on the generalized Poincare sphere is measured and adjusted by a perturbation strength.

To test the behavior of weight adaptation on dynamic channels, the linear ramp model was used to simulate dynamic channels of varying polarization rotation speeds. For the 7-core CC-MCF links, channel matrices for a link length of 13000 km were generated. The link length corresponds to the reach of a link with standard deviation of MDG $\sigma_g = 0.3$ dB per amplifier transmitting a 4-QAM signal and using ideal binary FEC with 20% overhead. For the SMF links, the link length was 11500 km. After generating the time-varying channel matrices, the LMS algorithm was used to adaptively determine the linear equalizer filter weights needed for SIC. Error-free FEC decoding was assumed during SIC. SIC was implemented using FDE with OS convolution with 50% overhead and computation blocks of size 512. The filter adaptation behavior of all implementations is similar for equal block sizes, but the CP method requires impractically long block sizes to keep transmission overhead of the CP low. Table I summarizes the loss in AIR resulting from increased MSE when adapting an equalizer to a dynamic channel. The LMS algorithm is able to track a 100 krad/s Stokes vector rotation with a 3% decrease in the AIR for $D = 2$ modes. Applying the same perturbation strength to a fiber with $D = 14$ modes results in a 400 krad/s rotation rate, which can be tracked by LMS if data-aided, but does not converge if decision-directed. If the rotation rate is decreased to 100 krad/s, however, the LMS algorithm is able to track channel changes using decision-directed feedback with AIR reduced by 12%. If the rotation rate is further decreased to 30 krad/s, the AIR reduction is less than 2%. A rotation rate of 30 krad/s in a fiber supporting $D = 14$ modes corresponds to a perturbation strength that induces a 4 krad/s rotation rate in a fiber supporting $D = 2$ modes, which is still higher than the maximum polarization rotation rate observed in the field measurements of [36].

V. CONCLUSION

SIC performed on a single-carrier signal using linear equalizers can mitigate the capacity penalty of MMSE detection in frequency-selective links with MDG with 1.8 to 2.6 times the complexity of MMSE detection, depending on the SIC implementation. An SDM fiber link exploiting the channel's full modal and frequency diversity does not suffer significant outage capacity reduction or LE penalty. However, to achieve the capacity associated with optimal detection using SIC, system architectures that perform FEC coding across transmit modes is required. Architectures that do not perform coding across modes have reduced modal diversity, incurring a penalty in AIR, and decreasing the advantage of SIC over MMSE detection. In a 7-core MC-MCF link with a standard deviation of MDG $\sigma_g = 0.3$ dB per amplifier, SIC can achieve 19% higher outage capacity than MMSE detection after 10000 km of transmission, but this reduces to 11% if data streams are coded individually.

When single-carrier transmission and linear equalizers are used to implement SIC in SMF systems, artificial PMD provides more benefits as PDG or link length are increased. Using ideal FEC with 20% overhead, adding per-span PMD with DGD uniformly distributed between -30 and +30 ps between 50 km fiber spans can almost double the transmission reach of a link transmitting 4-QAM with a standard deviation of PDG $\sigma_g = 0.3$ dB per amplifier.

ACKNOWLEDGMENT

The authors would like to thank Darli Mello, Karthik Choutagunta, and Hrishikesh Srinivas for helpful discussions about optimal capacity and MMSE AIR distributions, and Etsushi Yamazaki for helpful discussions about equalization and FEC in submarine links.

REFERENCES

[1] N. K. Dutta and Q. Wang, *Semiconductor Optical Amplifiers*, 2nd ed. New Jersey, NJ, USA: World Scientific, Feb. 2006, ch. 6.

[2] J. Renaudier and A. Ghazisaeidi, "Scaling capacity growth of fiber-optic transmission systems using 100 nm ultra-wideband semiconductor optical amplifiers," *J. Lightw. Technol.*, vol. 37, no. 8, pp. 1831–1838, Apr. 2019.

[3] Z. Zhu, X. Li, and Y. Xi, "A polarization insensitive semiconductor optical amplifier," *IEEE Photon. Technol. Lett.*, vol. 28, no. 17, pp. 1831–1834, Sep. 2016.

[4] M. Felicetti, J. J. van der Tol, E. J. Geluk, D. Pustakhod, M. J. Wale, and M. K. Smit, "Integrated high-performance TE/TM converters for polarization independence in semiconductor optical amplifiers," *J. Lightw. Technol.*, vol. 33, no. 17, pp. 3584–3590, Sep. 2015.

[5] S. Jeurink and P. M. Krummrich, "Multimode EDFA designs with reduced MDG by considering spatially dependent saturation effects," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, USA, 2018, pp. 1–3.

[6] M. Wada, T. Sakamoto, T. Yamamoto, S. Aozasa, and K. Nakajima, "Low mode dependent gain few-mode EDFA with fiber based mode scrambler," in *Proc. 24th OptoElectron. Commun. Conf. Int. Conf. Photon. Switching Comput.*, Fukuoka, Japan, 2019, pp. 1–3.

[7] M. Shtaif, "Performance degradation in coherent polarization multiplexed systems as a result of polarization dependent loss," *Opt. Exp.*, vol. 16, no. 18, pp. 13918–13932, Sep. 2008.

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

[9] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," *Opt. Exp.*, vol. 19, no. 17, pp. 16680–16696, Aug. 2011.

[10] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.

[11] A. J. Lowery, S. Wang, and M. Premaratne, "Calculation of power limit due to fiber nonlinearity in optical OFDM systems," *Opt. Exp.*, vol. 15, no. 20, pp. 13282–13287, Oct. 2007.

[12] K. Shibahara *et al.*, "Dense SDM (12-core × 3-mode) transmission over 527 km with 33.2-ns mode-dispersion employing low-complexity parallel MIMO frequency-domain equalization," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 196–204, Jan. 2016.

[13] A. Lozano and C. Papadias, "Layered space-time receivers for frequency-selective wireless channels," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 65–73, Jan. 2002.

[14] E. Awwad, G. R.-B. Othman, and Y. Jaoun, "Space-time coding schemes for MDL-impaired mode-multiplexed fiber transmission systems," *J. Lightw. Technol.*, vol. 33, no. 24, pp. 5084–5094, Dec. 2015.

[15] K. Shibahara *et al.*, "Iterative unreplicated parallel interference canceler for MDL-tolerant dense SDM (12-core × 3-mode) transmission over 3000 km," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1560–1569, Mar. 2019.

[16] K. Shibahara, T. Mizuno, D. Lee, and Y. Miyamoto, "Advanced MIMO signal processing techniques enabling long-haul dense SDM transmissions," *J. Lightw. Technol.*, vol. 36, no. 2, pp. 336–348, Jan. 2018.

[17] K. Shibahara, T. Mizuno, and Y. Miyamoto, "Long-haul mode multiplexing transmission enhanced by interference cancellation techniques based on fast MIMO affine projection," *J. Lightw. Technol.*, vol. 38, no. 18, pp. 4969–4977, Sep. 2020.

[18] H. Srinivas, E. S. Chou, D. A. A. Mello, K. Choutagunta, and J. M. Kahn, "Impact and mitigation of polarization- or mode-dependent gain in ultra-long-haul systems," in *Proc. 22nd Int. Conf. Transp. Opt. Netw.*, Bari, Italy, 2020, pp. 1–4.

[19] D. A. A. Mello, H. Srinivas, K. Choutagunta, and J. M. Kahn, "Impact of polarization- and mode-dependent gain on the capacity of ultra-long-haul systems," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 303–318, Jan. 2020.

[20] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 2012.

[21] R. Dar, M. Shtaif, and M. Feder, "New bounds on the capacity of the nonlinear fiber-optic channel," *Opt. Lett.*, vol. 39, no. 2, pp. 398–401, Jan. 2014.

[22] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. Int. Symp. Signals, Syst., Electron.*, Pisa, Italy, 1998, pp. 295–300.

[23] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[24] A. Andrusier, E. Meron, M. Feder, and M. Shtaif, "Equalization performance in the presence of linear polarization impairments," in *Proc. Opt. Fiber Commun. Conf.*, Los Angeles, CA, USA, 2012, pp. 1–3.

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

[26] S. O. Arik and J. M. Kahn, "Diversity-multiplexing tradeoff in mode-division multiplexing," *Opt. Lett.*, vol. 39, no. 11, pp. 3258–3261, Jun. 2014.

[27] R. Ryf *et al.*, "Coupled-core transmission over 7-core fiber," in *Proc. Opt. Fiber Commun. Conf. Exhibit.*, San Diego, CA, USA, 2019, pp. 1–3.

[28] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall Signal Processing Series). Englewood Cliffs, NJ, USA: Prentice Hall, 1989.

[29] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[30] S. O. Arik, J. M. Kahn, and K.-P. Ho, "MIMO signal processing for mode-division multiplexing: An overview of channel models and signal processing architectures," *IEEE Signal Process. Mag.*, vol. 31, no. 2, pp. 25–34, Mar. 2014.

[31] L. Li, A. Neubauer, and A. Czylwik, "Analysis of an iterative layered space time receiver with imperfect feedback," in *Proc. IEEE Veh. Technol. Conf.*, San Francisco, CA, USA, 2011, pp. 1–5.

[32] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, Oct. 2015.

[33] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[34] "100 G forward error correction white paper," OIF, OIF-FEC-100G-01.0. Accessed: May 2010. [Online]. Available: https://www.oiforum.com/wp-content/uploads/2019/01/OIF_FEC_100G-01.0-1.pdf

[35] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sep. 1984.

[36] N. Brochier and J.-L. Barbey, "Field measurement of polarization fluctuation dynamics and related impact for 40 Gbit/s submarine systems," in *Proc. SubOpt.*, Yokohama, Japan, 2010, pp. 1–5.

[37] K. Choutagunta, I. Roberts, and J. M. Kahn, "Efficient quantification and simulation of modal dynamics in multimode fiber links," *J. Lightw. Technol.*, vol. 37, no. 8, pp. 1813–1825, Apr. 2019.

Elaine S. Chou received the B.S.E. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2016, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 2018 and 2021, respectively. Her research interests include optical fiber communications, modulation and coding techniques, and signal processing.

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