

Fiber Impairment Compensation Using Coherent Detection and Digital Signal Processing

Ezra M. Ip and Joseph M. Kahn, *Fellow, IEEE*

(Invited Paper)

Abstract—Next-generation optical fiber systems will employ coherent detection to improve power and spectral efficiency, and to facilitate flexible impairment compensation using digital signal processors (DSPs). In a fully digital coherent system, the electric fields at the input and the output of the channel are available to DSPs at the transmitter and the receiver, enabling the use of arbitrary impairment precompensation and postcompensation algorithms. Linear time-invariant (LTI) impairments such as chromatic dispersion and polarization-mode dispersion can be compensated by adaptive linear equalizers. Non-LTI impairments, such as laser phase noise and Kerr nonlinearity, can be compensated by channel inversion. All existing impairment compensation techniques ultimately approximate channel inversion for a subset of the channel effects. We provide a unified multiblock nonlinear model for the joint compensation of the impairments in fiber transmission. We show that commonly used techniques for overcoming different impairments, despite their different appearance, are often based on the same principles such as feedback and feedforward control, and time-versus-frequency-domain representations. We highlight equivalences between techniques, and show that the choice of algorithm depends on making tradeoffs.

Index Terms—Adaptive signal processing, optical fiber communication.

I. INTRODUCTION

COHERENT detection has emerged as one of the key technologies in the development of high-speed, high-spectral-efficiency, dynamically reconfigurable optical networks suitable for long-distance transmission. By recovering the electric field in the two fiber polarizations, a coherent receiver allows information symbols to be encoded in all the degrees of freedom available in a fiber, leading to improved power and spectral efficiency.

Interest in coherent optical communications began in the late 1980s. Carrier synchronization was an early challenge due to the high carrier-linewidth-to-symbol-rate ratio compared to wireless and digital subscriber line (DSL) systems. Using narrow

linewidth lasers and optimized phase-locked loops (PLLs), coherent 4-Gb/s binary phase shift keying (BPSK) was demonstrated in [1], while 310-Mb/s quadriphase shift keying (QPSK) was demonstrated in [2]. Development in coherent optical systems was stalled in the 1990s by the invention of the erbium-doped fiber amplifier (EDFA), which enabled repeaterless transmission over long-haul distances. The combined spectra of the C-band (1530–1570 nm) and L-band (1570–1610 nm) offer a bandwidth of 10 THz. But with the rapid growth of Internet traffic, driven by new applications, such as video and music sharing, this bandwidth is rapidly becoming fully utilized. Thus, there is renewed interest in high-spectral-efficiency transmission.

Another factor contributing to the resurgence of coherent systems is the recent advance in very large scale integration (VLSI), which has made digital compensation of fiber impairments at gigahertz baud rates feasible. Digital compensation can be done either at the transmitter prior to upconversion onto an optical carrier, or at the receiver after the optical signal has been downconverted to the electronic domain. In both cases, provided the baseband signal in the electronic domain is sampled above the Nyquist rate, the digitized signal has the full information of the analog electric field, enabling digital signal processing (DSP) compensation to have no loss in performance compared to analog impairment compensation performed in either the optical or electronic domain. DSP has the advantage that signals can be delayed, split, amplified, and manipulated in other manner without degradation in signal quality. DSP-based receivers are already ubiquitous in wireless and DSL systems operating at comparatively low data rates.

In recent years, research on coherent systems has progressed at a rapid rate. Real-time coherent detection of polarization multiplexed 40-Gb/s QPSK (PM-QPSK) was demonstrated in [3], while 100-Gb/s systems and beyond are currently being developed. As the baud rate, constellation size, and transmission distances are increased, DSP algorithms will become increasingly complex to overcome a diversity of optical phenomena that currently limit achievable capacity.

A key difference between optical fiber transmission and other media, such as wireless or DSL, is that signal propagation is nonlinear, due to the tight confinement of light leading to high electric field intensity. Although nonlinearity also arises in radio-frequency (RF) transmission due to amplifier saturation, this nonlinearity is localized and instantaneous, and can be pre-compensated effectively. In optical fiber, nonlinearity imposes

Manuscript received June 08, 2009; revised July 02, 2009. First published July 24, 2009; current version published February 03, 2010.

E. M. Ip is with NEC Labs America, Princeton, NJ 08540 USA (e-mail: ezra.ip@nec-labs.com).

J. M. Kahn is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (e-mail: jmk@ee.stanford.edu).

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

Digital Object Identifier 10.1109/JLT.2009.2028245

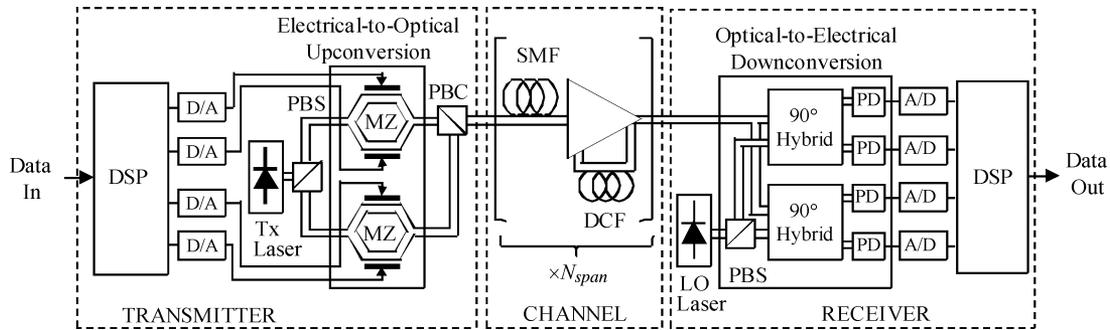


Fig. 1. Fully coherent optical system with digital signal processing at the transmitter and the receiver.

a fundamental limit on channel capacity. Digital compensation algorithms that can reduce deterministic nonlinear effects, and thereby increase capacity, are a subject of ongoing research.

Conversely, some channel impairments that degrade capacity in other channels are unimportant in fiber. For example, multipath fading is of critical importance in wireless transmission. In response, techniques such as orthogonal frequency-division multiplexing (OFDM) and multiple-input–multiple-output (MIMO) transmission have been developed. “Single-mode” fiber (SMF) actually supports two orthogonal polarization modes, making it a 2×2 MIMO system. Polarization-mode dispersion (PMD), while somewhat analogous to multipath propagation, is described by a channel matrix that is unitary in the absence of polarization-dependent loss (PDL). While in direct detection (DD) systems, PMD leads to fading, in coherent systems, PMD can be compensated losslessly by a linear equalizer. Nevertheless, coherent optical OFDM (CO-OFDM) has gained increasing interest, in part, because the transmitted spectrum can be dynamically optimized, which may be useful for dense wavelength-division multiplexing (DWDM) systems impacted by tight optical filtering. OFDM can also reduce the oversampling requirement compared to single-carrier (SC) transmission. Hence, techniques borrowed from other areas of communications often serve somewhat different purposes in optical fiber.

In this paper, we review DSP compensation algorithms for the compensation of various fiber impairments. Some techniques, such as linear equalization for combating chromatic dispersion (CD) and PMD, are already well understood and have been experimentally verified, while others, like nonlinear compensation, are emerging techniques with high computational cost, and will require further improvement in hardware for practical implementation.

The organization of this paper is as follows. In Section II, we present an analytical model of a fully digital coherent optical system that will serve as the basis for the sections to follow. We review signal propagation in fiber, and derive a linear, time-invariant (LTI) systems model that is accurate for low power and short transmission distances. In Section III, we review SC versus OFDM transmission. In Sections IV, we discuss linear equalizers for compensating LTI impairments, while in Sections V and VI, algorithms for combating laser phase noise and nonlinear impairments are presented. We will review algorithms

for both SC and OFDM, comparing their complexities and their scaling with respect to system parameters.

II. THEORY

A. System Model

A canonical model of a fully digital, coherent optical system is shown in Fig. 1. At the transmitter, a digital signal processor (DSP) converts input symbols into four digitized waveforms corresponding to the in-phase (I) and quadrature (Q) components of the two transmitted polarizations. These are converted to the analog domain by arbitrary waveform generators (AWGs), shown here as digital-to-analog (D/A) converters. Their outputs drive a pair of Mach–Zehnder (MZ) modulators for performing electrical-to-optical upconversion. The dually polarized modulated signal is then transmitted over a channel consisting of multiple spans of SMF, with amplification and dispersion compensation after each span. At the receiver, the optical signal is mixed with a local oscillator (LO) laser through an optical hybrid, followed by balanced detection. The optical front-end operations correspond to optical-to-electrical downconversion, and the four outputs are the baseband I and Q signals corresponding to the two received polarizations. These are sampled and digitized by analog-to-digital (A/D) converters, and then processed by a DSP, yielding the output symbols.

The system shown in Fig. 1 is sufficiently general to describe all coherent systems of interest in this paper. For example, the transmitter’s DSP can perform any arbitrary algorithm including dispersion and nonlinearity precompensation, and in the case of OFDM transmission, a subcarrier modulator may be included. Similarly, the receiver’s DSP can perform any arbitrary algorithm including dispersion and nonlinearity postcompensation, laser phase noise compensation, symbol synchronization, and in the case of OFDM, a subcarrier demodulator may also be included. Alternatively, if digital impairment compensation is only performed at one of the transmitter or the receiver, the DSP at the unused end may be omitted. In SC transmission, an additional stage of MZs may be inserted for pulse shaping [4].

TABLE I
PARAMETER DEFINITIONS FOR SINGLE AND MULTICHANNEL TRANSMISSION

	Single channel	Multichannel
No. of channels	$N_{ch} = 1$	$N_{ch} = 2$
Total received power	P_{rx}	P_{rx}
Noise psd for each channel (W/Hz)	N_0	N_0
Bits per symbol per channel	\bar{b}_{ch}	\bar{b}_{ch}
Total bits per multichannel symbol	$\bar{b} = \bar{b}_{ch}$	$\bar{b} = N_{ch} \bar{b}_{ch}$
SNR per symbol per channel*	$\gamma_s = \frac{P_{rx}}{N_0 B_W}$	$\gamma_s = \frac{P_{rx}}{N_{ch} N_0 B_W}$
SNR per bit	$\gamma_b = \gamma_s / \bar{b}_{ch}$	$\gamma_b = \gamma_s / \bar{b}_{ch}$
Shannon's limit per channel (b/s/Hz)	$C_{lim} = \log_2(1 + \gamma_s)$	$C_{lim} = \log_2(1 + \gamma_s)$

* For SC transmission, the effective bandwidth B_W is the symbol rate R_s . For OFDM, the signals of interest are the individual subcarriers. Thus, P_{rx} is the received power at the subchannel of interest, and $B_W = \Delta\omega/2\pi$ is the effective bandwidth (see Section III for definitions).

the received power needs to increase by the number of channels N_{ch} (two in the case of polarization multiplexing). Total capacity is then increased by N_{ch} , while the required SNR per bit is constant.

In Fig. 2, the impulse responses of the E-O and O-E converters are shown as $b(t)$ and $p(t)$, respectively. These are typically lowpass filters. $b(t)$ includes the frequency responses of the D/As, MZ modulators (including pulse carvers), and optical multiplexing elements, while $p(t)$ includes the responses of optical demultiplexing elements, the balanced photoreceivers, and the A/Ds.

Although the DSPs at the transmitter and the receiver can process signals at different rates, in this paper, we will assume a sampling interval of T . The definition of T for SC and OFDM transmission will be clarified in the following sections.

III. SINGLE-CARRIER AND MULTIPLE-CARRIER TRANSMISSION

In SC transmission without precompensation of impairments, the signal transmitted on each wavelength channel is

$$\mathbf{x}(t) = \sum_k \mathbf{x}_k b(t - kT_s) \quad (10)$$

where \mathbf{x}_k is the input symbol (a 2×1 complex-valued vector) transmitted at the k th symbol period, and $b(t)$ is the pulse shape. A disadvantage of SC transmission is that given the modulation format and pulse shape, the spectrum is fixed. The spectrum of RZ signals was studied in [8]. In a frequency selective channel, this is a suboptimal power allocation strategy. Consider the amplitude spectrum of a WDM channel shown in Fig. 3(a), where attenuation at high frequencies is caused by DWDM components such as interleavers and optical add/drop multiplexers (ROADMs). For fixed total transmitted power $\int_{-\infty}^{\infty} P(\omega) d\omega =$

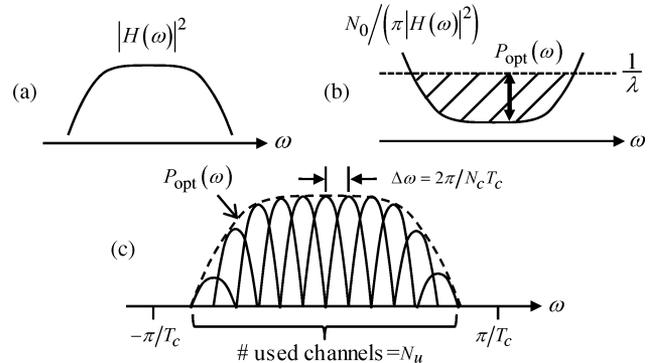


Fig. 3. (a) Amplitude response of channel, (b) optimum power allocation given by waterfilling, and (c) implementation of optimum spectrum using OFDM.

P_{tx} , the optimal power allocation strategy is given by waterfilling [Fig. 3(b)]. For a channel with no PDL and a noise psd given by (9), the optimum spectrum is

$$P_{opt}(\omega) = \left(\frac{1}{\lambda} - \frac{N_0}{\pi |H(\omega)|^2} \right)^+ \quad (11)$$

where $(x)^+$ denotes the larger of x or zero, and λ is chosen such that the shaded region $\int_{-\infty}^{\infty} P(\omega) d\omega$ in Fig. 3(b) is P_{tx} .

In multiple-carrier transmission, the available spectrum is divided into subcarrier “bins” [Fig. 3(c)]. By letting the frequency separation $\Delta\omega$ between subcarriers become infinitesimally small, and by modulating each of the N_c subcarrier with the correct power, the transmitted spectrum can be made arbitrarily close to (11). Different modulation formats and error-correction codes can be used on each subcarrier to approach capacity [9].

Orthogonal frequency-division multiplexing (OFDM) is a multiple-carrier format where demodulation and modulation onto subcarriers is performed digitally via the fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT). Compared to hardware-based frequency-division multiplexing (FDM) using unlocked oscillators, OFDM subcarriers are orthogonal over the duration of an “OFDM symbol” by design, so the signals transmitted in different subchannels can overlap in frequency and still be separable, enabling high spectral efficiency. Fig. 4(a) shows the operations performed by the OFDM modulators and demodulators. At the input, $\mathbf{X}_{1,k}$ and $\mathbf{X}_{2,k}$ are complex-valued vector of length N_c modulating the subcarriers at k th OFDM symbol period in the two signal polarizations. N_c is usually a power of two in order to exploit the computationally efficient IFFT algorithm, which converts the frequency-domain symbols into a time-domain signal. To mitigate against the channel’s ISI, a cyclic prefix of length N_{pre} is inserted, forming a vector of $N_c + N_{pre}$ “samples” that comprise an “OFDM symbol” [Fig. 4(b)]. The samples are transmitted serially at a rate of $1/T_c$, where T_c is the sample period. Provided $N_{pre}T_c$ exceeds the ISI duration, after propagation, the central N_c samples in the received OFDM symbol is given by the circular convolution between the transmitted OFDM symbol and the channel’s discrete-time impulse response. Thus, after stripping the prefix and taking the FFT, the received symbols can be

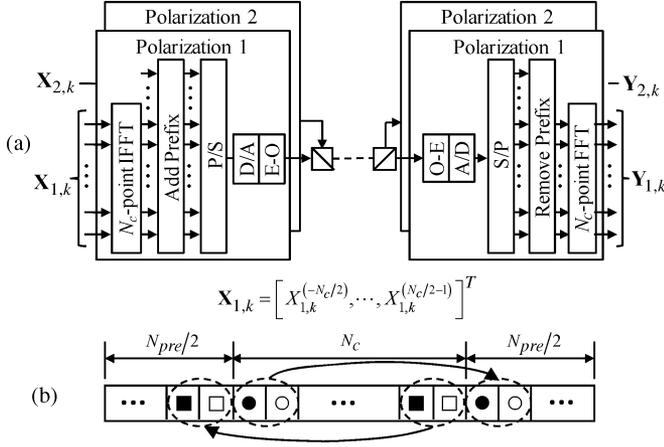


Fig. 4. (a) Dual-polarization OFDM modulator and demodulator, and (b) cyclic prefix insertion in an OFDM symbol.

recovered using a bank of “single-tap” equalizers which invert the phase distortions by CD and PMD at each subcarrier [10].

A dual-polarization OFDM signal has the form

$$\mathbf{x}(t) \approx \sum_{m=-N_c/2}^{N_c/2-1} \sum_k \mathbf{X}_k^{(m)} e^{jm\Delta\omega t} \text{rect}\left(\frac{t - kT_{\text{ofdm}}}{T_{\text{ofdm}}}\right) \quad (12)$$

where the approximation assumes the OFDM spectrum is entirely confined within the passband of the transmitter’s D/A, which has no amplitude or phase distortion. The separation between subcarriers is $\Delta\omega = 2\pi/N_c T_c$, and the OFDM symbol period is $T_{\text{ofdm}} = (N_c + N_{\text{pre}})T_c$.

Typically, only the center subcarriers N_u are modulated to prevent image band distortion by the transmitter’s D/A [Fig. 3(c)]. As the highest frequency component of the discrete-time OFDM signal is less than the Nyquist frequency π/T_c , having unused subcarriers is equivalent to oversampling by N_c/N_u . Depending on the characteristic of the D/A, $N_c/N_u \approx 5/4$ is generally sufficient for OFDM, whereas for SC, a higher oversampling rate of $M/K \approx 3/2$ is needed to the shallower rolloff of the SC spectrum compared to OFDM (see Section IV-A).

A disadvantage of OFDM is that insertion of the cyclic prefix requires the signaling rate to be increased to produce the same throughput, which leads to noise enhancement. From a frequency-domain perspective in (12), the cyclic prefix is equivalent to using only $\chi = N_{\text{pre}}/(N_c + N_{\text{pre}})$ of the available bandwidth within each subcarrier. χ is therefore the “bandwidth occupancy factor,” and $(1 - \chi)\Delta\omega$ is the effective width of the guard band between subcarriers. Assuming transmission is linear and no aliasing effects, OFDM requires the SNR at the receiver to be higher by $1/\chi$ to achieve the same bit error ratio (BER) as SC. Higher launch power makes OFDM more susceptible to nonlinear effects. Furthermore, for large N_u , the field of an OFDM signal becomes Gaussian, which has high

TABLE II
PARAMETER EQUIVALENCE BETWEEN SC AND OFDM SIGNALS

	Single Carrier	OFDM
Total no. of subcarriers	1	N_c
No. of used subcarriers	1	N_u
Prefix length	0	N_{pre}
Time unit	$R_s = 1/T_s$	$R_c = 1/T_c$
Bandwidth occupancy	$\chi = 1$	$\chi = \frac{N_c}{N_c + N_{\text{pre}}}$
Oversampling rate	M/K^*	N_c/N_u
Effective symbol rate	R_s	$R_{s,\text{eff}} = \frac{N_u}{N_c + N_{\text{pre}}} R_c = \frac{N_u}{N_c} \chi R_c$
ISI cancellation	Equalizer length N	Prefix length N_{pre}

* Defined in Section IV-A.

peak-to-average power ratio (PAPR). This further enhances nonlinearity. In Section VI, we will see that nonlinear tolerance can be improved using dispersion unmanaged transmission. In this regime, there is little performance difference in performance between SC and OFDM.

Table II shows the parameter equivalences between SC and OFDM transmission. In SC, the fundamental time unit is the symbol interval T_s , whereas the equivalent time unit in OFDM is the sample interval $T_c = T_{\text{ofdm}}/(N_c + N_{\text{pre}})$. Since OFDM has N_u subcarriers signaling at a rate of $R_{\text{ofdm}} = 1/T_{\text{ofdm}}$, we define the “effective symbol rate” of OFDM to be $R_{s,\text{eff}} = N_u R_{\text{ofdm}}$. In Section IV, we will show that $R_{s,\text{eff}}$ is a useful parameter for computing the prefix length requirement, as it enables the formulas derived for SC to be reused. We note that SC transmission is a special case of OFDM where $N_u = N_c = 1$ and $N_{\text{pre}} = 0$. However, whereas SC transmission typically often uses a dedicated pulse carver to control the pulse shape $b(t)$, in OFDM described above, rectangular pulses $b(t) = \text{rect}(t/T_{\text{ofdm}})$ of width T_{ofdm} are used for all subcarriers.¹

IV. LINEAR EQUALIZATION

Since CD and PMD are LTI impairments described by unitary matrices in the frequency domain, they can be compensated losslessly by a linear equalizer. Depending on the modulation format (SC or OFDM), the ISI duration of the channel and the rate of its variation, different equalizer realizations can be used. In this section, we review time-domain equalization, frequency-domain equalization, and hybrid structures, highlighting their advantages and disadvantages, and discuss their computational requirements. We will review common adaptive algorithms used with these equalizer realizations.

¹It is possible to perform pulse shaping in OFDM. See [11].

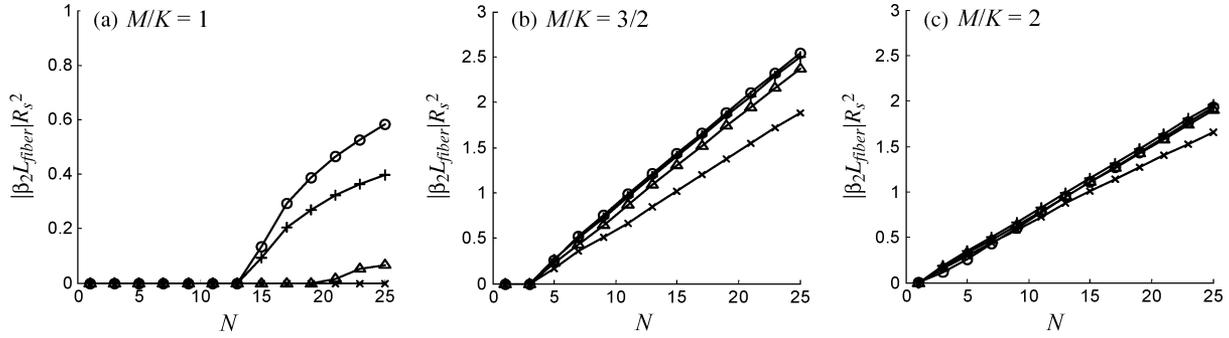


Fig. 5. Normalized residual CD versus equalizer length for oversampling rates of (a) 1, (b) 3/2, and (c) 2. For each plot, the curves denote: o: NRZ, x: 33% RZ, Δ : 50% RZ, and +: 67% RZ.

A. Time-Domain Equalizer (TDE)

In SC transmission, we sample each polarization of the received signal at a rate of $1/T = (M/K)(1/T_s)$, where M/K is a rational oversampling ratio

$$y_{i,k} \triangleq y(kT) = \sum_n \sum_{j=1}^2 x_{j,n} q_{ij}(kT - nT_s) + n'_i(kT). \quad (13)$$

$q_{ij}(t) = p(t) \otimes h_{ij}(t) \otimes b(t)$ is the received pulse shape between the j th input and the i th output polarizations pair, as defined in Fig. 2, and $n'_i(t) = p(t) \otimes n_i(t)$ is noise at the i th output polarization. If successive transmitted symbols, received samples, and noise samples are written as vectors, and the two polarization signals are padded together, we note that (13) can be expressed as an equation of the form $\mathbf{Y} = \mathbf{Q}\mathbf{X} + \mathbf{N}'$, where \mathbf{Q} is a matrix with a band diagonal structure, where the width of the band corresponds to the ISI duration. To estimate the k th transmitted symbol, we select the N nearest neighbors to sample $\lfloor kM/K \rfloor$ which lies closest in time to the peak of the k th symbol. For each polarization, we form the vector

$$\mathbf{y}_{i,k} \left[y_{i, \lfloor \frac{kM}{K} \rfloor + \lfloor \frac{N}{2} \rfloor}, y_{i, \lfloor \frac{kM}{K} \rfloor + \lfloor \frac{N}{2} \rfloor - 1}, \dots, y_{i, \lfloor \frac{kM}{K} \rfloor - \lfloor \frac{N}{2} \rfloor + 1} \right]^T. \quad (14)$$

Linear equalization is then given by a matrix multiplication

$$\hat{\mathbf{x}}_k = \begin{bmatrix} \hat{\mathbf{x}}_{1,k} \\ \hat{\mathbf{x}}_{2,k} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_{11}^T & \mathbf{w}_{21}^T \\ \mathbf{w}_{12}^T & \mathbf{w}_{22}^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1,k} \\ \mathbf{y}_{2,k} \end{bmatrix} = \mathbf{w}^T \mathbf{y}_k \quad (15)$$

which corresponds to a finite impulse response (FIR) filter per input–output polarization pair. The equalizer \mathbf{w} is a complex-valued matrix of size $2N \times 2$, and the partitions $\mathbf{w}_{ij} = [\mathbf{w}_{ij, -\lfloor N/2 \rfloor}, \mathbf{w}_{ij, -\lfloor N/2 \rfloor + 1}, \dots, \mathbf{w}_{ij, \lfloor N/2 \rfloor - 1}]^T$ are column vectors of length N . A direct implementation of (15) thus requires $4N$ complex multiplications by symbol. The optimum Wiener filter which minimizes mean squared error (MSE) $E[\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2]$ is given by [12]

$$\mathbf{w}_{\text{opt}} = \mathbf{A}^{-1} \boldsymbol{\alpha} \quad (16)$$

where $\mathbf{A} = E[\mathbf{y}_k^* \mathbf{y}_k^T]$ and $\boldsymbol{\alpha} = E[\mathbf{y}_k^* \mathbf{x}_k^T]$. We can further partition these autocorrelation and cross-correlation matrices into submatrices $\mathbf{A}_{ij} = E[\mathbf{y}_{i,k}^* \mathbf{y}_{j,k}^T]$ of size $N \times N$, and vectors

$\boldsymbol{\alpha}_{ij} = E[\mathbf{y}_{i,k}^* \mathbf{x}_{j,k}]$ of length N . Let $A_{ij,lm}$ be the element at the intersection of the l th row and m th column of \mathbf{A}_{ij} , and let $\alpha_{ij,l}$ be the element at the l th row of $\boldsymbol{\alpha}_{ij}$. Assuming the transmitted symbols are independent and identically distributed (i.i.d.), it can be shown that [13]

$$\begin{aligned} A_{ij,lm} &= P_{rx} \sum_n \sum_{s'=1}^2 q_{is'}^*(rT/K - lT - nT_s) \\ &\quad \times q_{js'}(rT/K - mT - nT_s) \\ &\quad + N_0 \int_{-\infty}^{+\infty} p^*(t - lT) p(t - mT) dt, \end{aligned} \quad \text{for } -\lfloor N/2 \rfloor \leq l, m \leq \lfloor N/2 \rfloor - 1 \quad (17)$$

and

$$\alpha_{ij,l} = P_{rx} q_{ij}^*(rT/K - lT), \quad \text{for } -\lfloor N/2 \rfloor \leq l \leq \lfloor N/2 \rfloor - 1 \quad (18)$$

where $r = kM \bmod K$. When a noninteger oversampling rate is used ($K > 1$), there are K unique values of r corresponding to all the possible sampling phases relative to the symbol peak. Thus, the equalizer in fact comprises a bank of K matrices whose coefficients are the K unique solutions of (16). These are used successively by the receiver to recover the transmitted symbols, at the symbol rate of $1/T_s$.

1) *Equalizer Length and Oversampling*: It was shown in [13] that when the receiver A/D is well modeled by a fifth-order Butterworth filter response with a 3-dB cutoff frequency equal to 40% of the symbol rate, a minimum oversampling rate of 3/2 is necessary to mitigate against sampling time offset, and to enable the compensation of an arbitrary amount of residual CD and PMD by increasing the equalizer length. When an insufficient oversampling rate is used, aliasing by the A/D causes signal cancellation in the frequency domain. The resulting noise enhancement leads to unacceptable power penalty [14], [15]. Fig. 5 shows the equalizer length required to overcome residual CD with 2-dB power penalty. For oversampling rates of 3/2 or higher, the residual CD compensated increases linearly with N , but for symbol-rate sampling, the maximum tolerable residual CD only reaches an asymptote, as the combination of phase distortion and aliasing causes signal cancellation.

The slope of Fig. 5 describes the increase in computational complexity as a function of residual dispersion. A channel

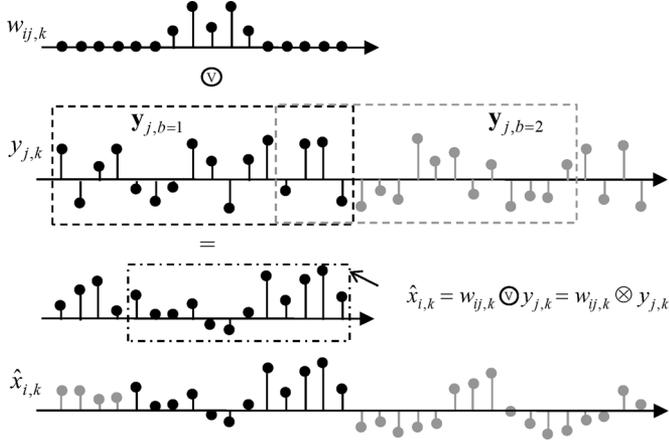


Fig. 6. Frequency-domain implementation of convolution using overlap-and-save.

whose residual dispersion is $(\beta_2 L)_{\text{res}}$ has an ISI duration of N_{cd} symbols due to CD

$$N_{\text{cd}} \approx 2\pi |(\beta_2 L)_{\text{res}}| R_s^2. \quad (19)$$

Similarly, the worst case ISI arising from PMD depends on the maximum DGD $\Delta\tau_{\text{max}}$, and spans N_{pmd} symbols

$$N_{\text{pmd}} \approx \Delta\tau_{\text{max}} R_s \approx 3\overline{\Delta\tau} R_s. \quad (20)$$

Finally, all other LTI distortions, including optical and electrical filtering, may contribute significant ISI over N_{ex} symbols. The required equalizer length is the sum of the individual ISI durations multiplied by the oversampling rate

$$N > (N_{\text{cd}} + N_{\text{pmd}} + N_{\text{ex}})(M/K). \quad (21)$$

Unless the dispersion map is carefully controlled to ensure near-zero residual CD at the receiver, the greatest contribution to (21) usually arises from N_{cd} .

2) *Computational Requirement:* Since the matrix multiplication operation in (15) is evaluated every symbol, the complexity of direct implementation is $4NR_s$ complex multiplications per symbol. A benefit of direct implementation is its low latency, corresponding to one complex multiplication and $\log_2(N)$ additions.

B. Frequency-Domain Equalizer (FDE)

For large N , FIR filtering is more efficiently implemented in the frequency domain using the overlap-and-add or overlap-and-save methods [16]. The overlap-and-save method is shown in Fig. 6 considering only one polarization. The received signal at the i th polarization is partitioned into overlapping blocks $\mathbf{y}_{j,b} = \{y_{i,k}\}_{k=bB-N+1}^{(b+1)B-1}$ of length $V = N + B - 1 = 2^v$ that is an integer power of two. We perform a V -point circular convolution between $\mathbf{y}_{j,b}$ and the equalizer \mathbf{w}_{ij} by taking the FFT of the two vectors, multiply them together, and then taking the IFFT. The last B samples of the circular convolution are the same as the linear convolution. These output samples are saved. The next input block $\mathbf{y}_{j,b+1}$ is obtained by advancing B samples, and the process is repeated.

1) *Computational Requirement:* A V -point FFT or IFFT requires $(V/2)\log_2 V$ complex multiplications to compute, while multiplying $\text{FFT}\{\mathbf{w}_{ij}\}$ by $\text{FFT}\{\mathbf{y}_{j,b}\}$ requires a further V complex multiplications. For dually polarized signals, (15) correspond to four FIR filters, while the number of Fourier transforms is doubled. Since each block has to be processed within B symbols, the number of complex multiplications performed per symbol is therefore

$$\min_{B:N+B-1=2^v} \left(\frac{1}{B} (N+B-1) (2\log_2(N+B-1)+4) \right) R_s. \quad (22)$$

For a given N , there exists an optimal block size that minimizes (22). It can be shown that the complexity per symbol scales as $O(\log_2 N)$, whereas the complexity for direct implementation scales as $O(N)$. A disadvantage of a frequency-domain implementation is its high latency, which is around $2\log_2 V + 1$ complex multiplications and additions, caused by the FFT and IFFT operations.

Fig. 7(b) shows a block diagram of the FDE for SC. The equalizer coefficients $\mathbf{W} = \text{FFT}\{\mathbf{w}\}$ are stored in the frequency domain. If the LTI model of the channel is valid, the IFFT operation can be rearranged to the transmitter side [Fig. 7(c)], and the inputs and outputs of this system are frequency-domain symbols modulating different subcarriers. Thus, frequency-domain equalization for SC and OFDM are analogous. The processing block of length V in the FDE for SC is exactly analogous to an OFDM symbol. Table III summarizes their equivalent parameters. In OFDM, the need to overlap successive blocks is avoided by using a cyclic prefix whose duration $N_{\text{pre}}T_c$ should exceed the ISI duration. The condition for this requirement is exactly the same as given by (19)–(21), provided the equivalent OFDM parameters are substituted for SC parameters as per Tables II and III. For a given target data rate, oversampling rate and bandwidth occupancy factor, an estimate of $R_{s,\text{eff}}$ can be computed, and is substituted into (19) and (20). The prefix length is then given by (21), with N_{pre} being OFDM's equivalent of the SC's linear equalizer length N .

With no aliasing between the received OFDM symbols, it can be shown that the input and output signals are related by

$$\mathbf{Y}(\omega_m) = \mathbf{Q}(\omega_m)\mathbf{X}(\omega_m) + \mathbf{N}'(\omega_m) \quad (23)$$

where $\omega_m = m\Delta\omega$ is the frequency of the m th subcarrier, where $-[N_u/2] \leq m \leq [N_u/2] - 1$, and $\mathbf{Q}(\omega)$ is the frequency response of the channel given by

$$p(t) \otimes \mathbf{H}(t) \otimes b(t) = \mathbf{Q}(t) \xrightarrow{\mathcal{F}} \mathbf{Q}(\omega) = P(\omega)\mathbf{H}(\omega)B(\omega). \quad (24)$$

The FDE for OFDM is a bank of 2×2 matrices multiplying the received symbol at each subcarrier

$$\hat{\mathbf{X}}(\omega_m) = \mathbf{W}(\omega_m)\mathbf{Y}(\omega_m). \quad (25)$$

It can be shown that the optimum Wiener filter is given by

$$\mathbf{W}_{\text{opt}} = \mathbf{S}_{\text{XX}}\mathbf{Q}^H \left(\mathbf{Q}\mathbf{S}_{\text{XX}}\mathbf{Q}^H + \frac{1}{\chi}\mathbf{P}\mathbf{S}_{\text{NN}}\mathbf{P}^H \right)^{-1} \quad (26)$$

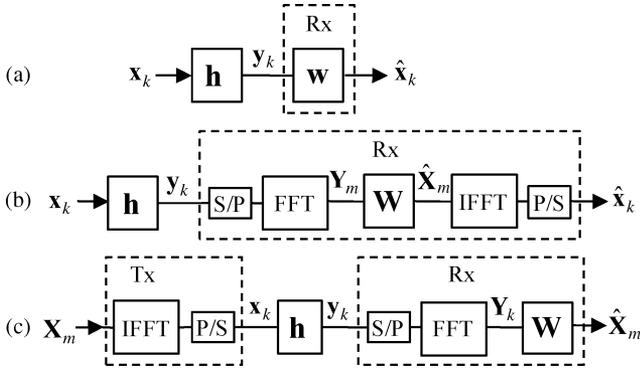


Fig. 7. (a) SC transmission with TDE, (b) SC transmission with FDE, and (c) OFDM transmission with FDE.

TABLE III
EQUIVALENCE PARAMETERS BETWEEN FDE FOR SC AND OFDM

	Single Carrier	OFDM
ISI mitigation requirement	Equalizer length N	Prefix length N_{pre}
Block size	B	No. of subcarriers N_c
Overlap between adjacent blocks	$B-1$	0
FFT size (Power of two)	$V = N + B - 1$	No. of subcarriers N_c

where $\mathbf{S}_{XX}(\omega_m) = E[\mathbf{X}(\omega_m)\mathbf{X}^H(\omega_m)]$ is the signal psd matrix, and $\mathbf{S}_{NN}(\omega_m) = E[\mathbf{N}(\omega_m)\mathbf{N}^H(\omega_m)]$ is the noise psd matrix. Note that the noise has been scaled by the inverse of χ due to the power penalty of the cyclic prefix [11]. If the noises in orthogonal polarizations are uncorrelated, polarization-multiplexed transmission is used, the data modulating the two polarizations have equal power and are uncorrelated, and the channel's frequency response is flat over all the subcarriers, (26) simplifies to

$$\mathbf{W} = \frac{\gamma_s}{\gamma_s + 1} \mathbf{Q}^H \quad (27)$$

which is a unitary filter that exactly inverts the phase distortions of CD, PMD, and other optical or electrical filtering effects.

2) *Computational Requirement*: The complexity of OFDM is analogous to that of SC using a FDE, except for the IFFT which is performed by the transmitter. Since the FFT and IFFT sizes are N_c , and the number of used subcarriers requiring equalization is N_u , the transmitter and receiver complexities are

$$\left(\frac{1}{N_u} N_c \log_2 N_c\right) R_s \quad (28)$$

and

$$\left(\frac{1}{N_u} N_c \log_2 N_c + 4\right) R_s. \quad (29)$$

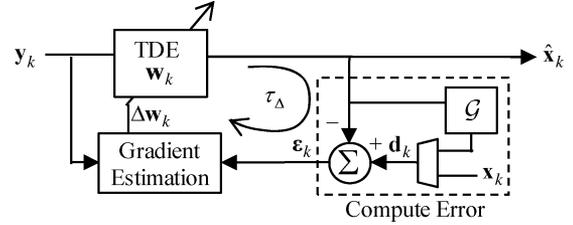


Fig. 8. Adaptive TDE.

TABLE IV
COST FUNCTIONS AND DESIRED SIGNALS OF COMMONLY USED ADAPTIVE ALGORITHMS

	Cost Function (J)	Desired Signal (\mathbf{d}_k)
Decision-Aided (DA)	$E\left[\ \mathbf{x}_k - \hat{\mathbf{x}}_k\ ^2\right]$	\mathbf{x}_k
Decision-Directed (DD)	$E\left[\ \llbracket \hat{\mathbf{x}}_k \rrbracket_D - \hat{\mathbf{x}}_k\ ^2\right]$	$\llbracket \hat{\mathbf{x}}_k \rrbracket_D$
Constant Modulus (CMA)	$E\left[\ \hat{\mathbf{x}}_k\ ^2 - \mathbf{R}^{(2)}\ ^2\right]$	$\begin{bmatrix} \hat{x}_{1,k} \left(1 + R_1^{(2)} - \hat{x}_{1,k} ^2\right) \\ \hat{x}_{2,k} \left(1 + R_2^{(2)} - \hat{x}_{2,k} ^2\right) \end{bmatrix}$
Godard	$E\left[\ \hat{\mathbf{x}}_k\ ^p - \mathbf{R}^{(p)}\ ^2\right]$	$\begin{bmatrix} \hat{x}_{1,k} \left(1 + R_1^{(p)}\right) \hat{x}_{1,k} ^{p-2} - \hat{x}_{1,k} ^{2p-2} \\ \hat{x}_{2,k} \left(1 + R_2^{(p)}\right) \hat{x}_{2,k} ^{p-2} - \hat{x}_{2,k} ^{2p-2} \end{bmatrix}$

C. Adaptive TDE

In practice, the channel is time varying due to mechanical vibration, changes in temperature, and other environmental conditions. In particular, PMD can fluctuate on a time scale of a millisecond. Hence, the linear equalizer needs to be adaptive. Adaptive equalizers have been studied in [14] and [17].

The canonical model of an adaptive equalizer is shown in Fig. 8. The received samples are processed by the linear equalizer, whose output $\hat{\mathbf{x}}_k$ is subtracted from the desired signal to obtain an error signal

$$\boldsymbol{\varepsilon}_k = \mathbf{d}_k - \hat{\mathbf{x}}_k. \quad (30)$$

1) *Error Signal*: The receiver uses this error to adjust the equalizer taps through a gradient estimation block which computes the coefficient adjustment to minimize a cost function J . Depending on the adaptive algorithm used, the desired signal is either the training symbols \mathbf{x}_k , or the equalizer output passed through a nonlinear function $\mathcal{G}\{\hat{\mathbf{x}}_k\}$. Table IV shows cost functions and desired signals for some commonly used adaptive algorithms.

If training symbols are periodically inserted, data-aided (DA) training can be used, where the error signal $\boldsymbol{\varepsilon}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ is the difference between the known training symbols and the equalizer output.

Training symbols reduce throughput, however. There may also be synchronization issues when the system is first started. Thus, in many applications, training symbols are omitted. During normal transmission, for example, the receiver can

operate in decision-directed (DD) mode where symbol decisions made are assumed to be correct, and the error signal $\boldsymbol{\epsilon}_k = [\hat{\mathbf{x}}_k]_D - \hat{\mathbf{x}}_k$ is the difference between the decision and the equalizer output. Provided the symbol-error rate is low, DD operation is exactly analogous to DA. However, a low symbol-error rate is not always guaranteed, particularly when the receiver is first started, or after a major interruption such as the carrier synchronizer becoming unlocked.

In the absence of training symbols, “blind” adaptive algorithms can be used to perform an initial training of the equalizer. When the symbol error rate becomes sufficiently low—equivalent to the “eye” becoming open—the system can switch to DD mode. Blind adaptive algorithms exploit known properties of the transmitted signal. In phase-modulated constellations, for example, the signal points lie on a circle, enabling the use of radius-directed adaptive algorithms. Godard introduced a family of such algorithms based on minimizing the cost function $J = E[|\hat{x}_k|^p - R^{(p)}]^2$, where p is a positive integer and $R^{(p)} = E[|x_k|^{2p}]/E[|x_k|^p]$ is the target radius [18]. The case of $p = 2$ is most commonly used, and is known as the “constant modulus algorithm” (CMA) [19]. For dual-polarization signals, the CMA can be generalized to minimize the cost function $J_i = E[|\hat{x}_{i,k}|^p - R_i^{(p)}]^2$ for each polarization. Since $R_i^{(p)} = E[|x_{i,k}|^{2p}]/E[|x_{i,k}|^p]$ can be computed for any arbitrary constellation, the Godard algorithm works also for nonconstant modulus constellations such as QAM—albeit with loss of performance for densely packed constellations. Other algorithms that are suitable for QAM have also been studied, and they include the multimodulus algorithm (MMA) [20] and the reduced constellation algorithm (RCA) [21].

2) *Coefficient Update*: The equalizer coefficients are then adjusted according to

$$\mathbf{w}^{(m+1)} = \mathbf{w}^{(m)} - \mu \nabla_{\mathbf{w}} J \quad (31)$$

where μ is the step size. The desired signals in Table IV have been defined such that

$$\nabla_{\mathbf{w}} J = -2E[\mathbf{y}_k^* \boldsymbol{\epsilon}_k^T] = -2E[\mathbf{y}_k^* (\mathbf{d}_k - \hat{\mathbf{x}}_k)^T]. \quad (32)$$

Thus, when the equalizer is at its optimum setting, the error is orthogonal to the received signal. For noninteger oversampling, we recall that there are K sets of coefficients for each of the possible sampling phases. The superscript in (31) denotes the m th update of that equalizer. For example, when an oversampling rate of 3/2 is employed, different coefficients are used for odd and even symbols, so these are adapted independently. By dissociating the superscript m from the symbol number k , it is assumed that the equalizer coefficients may not be updated every symbol. This is particularly important for optical systems where the latency is much greater than a symbol period, so the TDE coefficients may not be adapted every symbol.

When the equalizer coefficients are adapted using the exact gradient function in (32), the coefficient update method is known as “steepest descent,” and the rate of convergence is related to the step size and to the spread of eigenvalues of the autocorrelation matrix $\mathbf{A} = E[\mathbf{y}_k \mathbf{y}_k^H]$. The use of larger step size results in faster convergence, but leads to larger “misadjustment” in the presence of noise [17], and can even

cause the coefficients to diverge. For the DA algorithm, stability requires $0 < \mu < 1/\lambda_{\max}$, where λ_{\max} is the largest eigenvalue of \mathbf{A} . By contrast, the time constant for convergence $N_{\text{iter}} = 1/(\mu\lambda_{\min})$ is limited by the smallest eigenvalue λ_{\min} of \mathbf{A} corresponding to the slowest mode.

Since \mathbf{A} is an autocorrelation matrix, its eigenvalues are related to the channel’s frequency response. As CD and PMD are unitary transformations, the frequency response is flat except at the highest frequencies which may be attenuated by optical filters. Over the frequencies of interest, the spread of eigenvalues is small, and we have $\lambda_{\min} \approx \lambda_{\max} \approx P_{rx}$.

3) *Least Mean Square*: In practice, the exact gradient function in (32) is difficult to estimate. Many samples may be needed to gain sufficient knowledge of $\nabla_{\mathbf{w}} J$ before an update is possible. Such a delay may lead to intolerable adaptation speed. In the least mean square (LMS) algorithm, a single-shot estimate is used for the gradient

$$\nabla_{\mathbf{w}} J \approx -2\mathbf{y}_k^* \boldsymbol{\epsilon}_k^T \quad (33)$$

where the error $\boldsymbol{\epsilon}_k$ is defined by (30) and Table IV. As $\nabla_{\mathbf{w}} J$ is available after processing each symbol, the equalizer can adapt as rapidly as processing latency allows. The time constant is therefore N_{iter} times the latency of the feedback loop (Fig. 8), which we define as τ_{Δ} . The cost of computing the gradient in (33) is $4N$ complex multiplications. If the equalizer is adjusted every symbol, the computational cost of adaptation is 100%.

4) *Recursive Least Squares*: A class of recursive least squares (RLS) algorithms achieves faster convergence than LMS by computing a more accurate estimate of the gradient, which is recursively updated after each symbol. RLS was studied in [17], while a modified RLS algorithm suitable for use with CMA was studied in [22]. RLS suffers from high complexity due to the need for matrix inversion, and may not be suitable for implementation at the baud rates of optical systems.

D. Adaptive FDE

To reduce computational cost, an adaptive FDE can be implemented by inserting FFTs and IFFTs into Fig. 8. This is shown in Fig. 9(a). The coefficients of an FDE are frequency-domain values. However, both the received signal and the error are time-domain values, as the latter is obtained by subtracting the equalizer output from a time-domain desired signal \mathbf{d}_k . To adjust the equalizer coefficients, the error must therefore also be transformed to the frequency domain.

Adaptive FDEs were studied in [23], and were applied to coherent optical systems in [24]. It can be shown that the operations performed by the adaptive FDE are completely analogous to an adaptive TDE, with the same performance and step size requirement [17]. Thus, for the DA, DD, and blind algorithms, the error signal can be found using the formulas shown in Table IV.

In real-time implementation, it is not possible to wait for the entire signal to be received before computing the FFTs/IFFTs. The received signal is therefore parallelized and processed in blocks. The most commonly used block size is $2N$, where N is the length of the equalizer (which is padded with zero to ensure N is a power of two). In frequency-domain LMS algorithm (FLMS), the gradient estimation block contains an additional FFT and IFFT, giving a total complexity of five Fourier

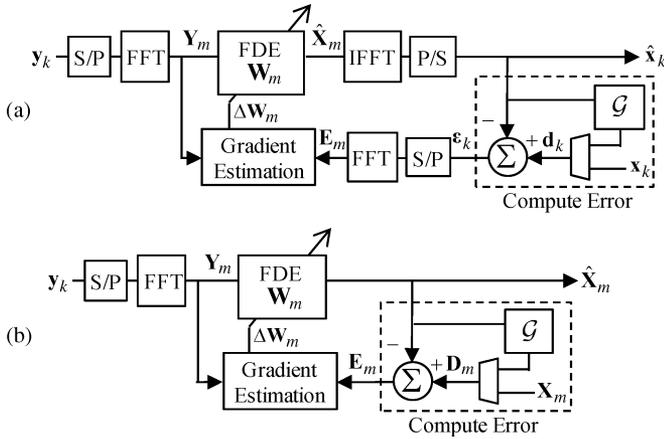


Fig. 9. Adaptive FDE for (a) SC and (b) OFDM.

transforms. Compared to the nonadaptive FDE, the complexity per symbol is increased by $\sim 150\%$ if the equalizer is adjusted after every block. A computationally simpler “unconstrained gradient” algorithm can also be used with only 50% complexity increase over nonadaptive FDE. But this algorithm has worse performance than standard FLMS [25].

The major disadvantage of an adaptive FDE comes from feedback delay, due to the presence of FFTs and IFFTs in the feedback loop of Fig. 9(a). As the coefficients cannot be updated faster than once every feedback delay, the time constant of the adaptive FDE $N_{\text{iter}}\tau_{\Delta(\text{FDE})}$ is much longer than the time constant $N_{\text{iter}}\tau_{\Delta(\text{TDE})}$ of an adaptive TDE, which has no FFTs.

E. Adaptive Multidelay Block FDE

It is possible to tradeoff the fast adaptation of a TDE against the low computational cost of an FDE by using a multidelay-block frequency-domain (MDF) equalizer [26]. The MDF parallelizes the input signal into block sizes of $2N/B$, where B is the number of processing blocks (Fig. 10). Each of the blocks contains an adaptive FDE. As the FFT/IFFT sizes are $2N/B$, the number of butterfly stages in each FFT is reduced by $\log_2 B$, resulting in shorter latency. However, computational cost is increased, as the FFT/IFFT operations have to be performed at B times the rate of a full FDE. Both the FDE and the TDE are in fact limiting cases of an MDF when the number of blocks are $B = 1$ and $B = 2N$, respectively. Table V summarizes the latency and computational requirement of various adaptive equalizer structures.

F. Serial TDE/FDE

In most fibers, CD is relatively static, while PMD is time varying. This enables the use of a hybrid equalizer structure shown in Fig. 11 to improve the performance of SC systems [27]. For CD compensation, a nonadaptive FDE is computationally efficient, while for PMD compensation, an adaptive TDE has the least latency. This structure is the most widely used for real-time coherent optical systems. The length of the TDE is typically between 7 and 13 taps, while the length of the FDE depends on the residual dispersion in the system. From (19) and (21), it is observed that each equalizer tap compensates $1/2\pi R_s^2 (MK)$ of residual CD. Assuming a baud rate of

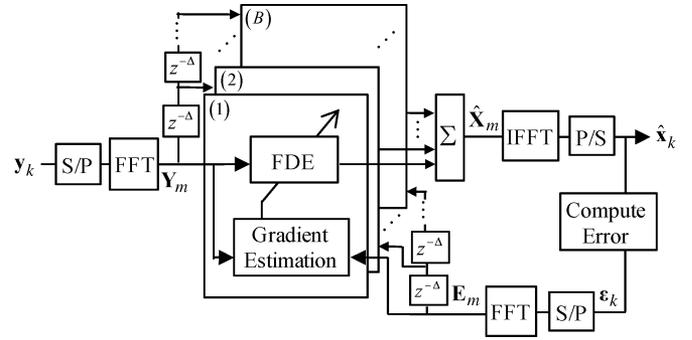


Fig. 10. Adaptive multidelay block FDE.

TABLE V
LATENCY AND COMPUTATIONAL COST SCALING WITH EQUALIZER LENGTH FOR TDE, FDE, AND MDF

	Processing Requirement (Multiplications per symbol)	Feedback Latency (# of Multiplications)
Time-Domain Equalizer (TDE)	$O(N)$	$O(1)$
Frequency-Domain Equalizer (FDE)	$O(\log_2 N)$	$O(\log_2 N)$
Multidelay block FDE (MDF)	$O\left(\log_2 \frac{N}{B}\right) + O(B)$	$O\left(\log_2 \frac{N}{B}\right) + O(1)$

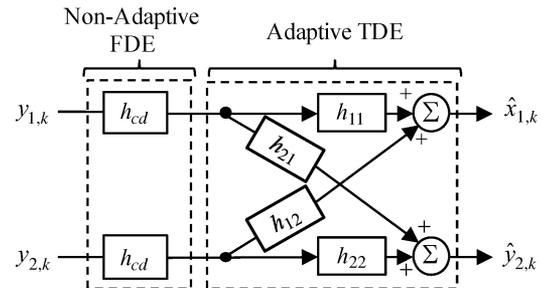


Fig. 11. Hybrid FDE-TDE structure for compensating CD and PMD.

25 GHz and an oversampling rate of two, this corresponds to 100 ps/nm per tap at a carrier wavelength of 1550 nm. Equalization of very large residual dispersion arising from dispersion unmanaged transmission has been demonstrated. For example, Savory *et al.* [28] successfully detected 42.8-Gb/s QPSK transmitted over 6400 km of SMF without DCF, while Charlet *et al.* [29] detected 112-Gb/s QPSK transmitted over 7040-km large effective area fiber (LEAF) without DCF, using a 1500 tap equalizer.

G. OFDM

OFDM can simultaneously exploit the computational advantage of an FDE without the latency of FFTs in the feedback path by transmitting symbols in the frequency domain. An adaptive FDE for OFDM is shown in Fig. 9(b). To compute the error signal, the formulas in Table IV can be used by replacing the time-domain signals \mathbf{d}_k , \mathbf{x}_k , and $\hat{\mathbf{x}}_k$ with their frequency-domain equivalents \mathbf{D}_k , \mathbf{X}_k , and $\hat{\mathbf{X}}_k$ in Fig. 9(b). As the desired signal is computed in the same (frequency) domain as the equalizer coefficients, no transformation is required to obtain

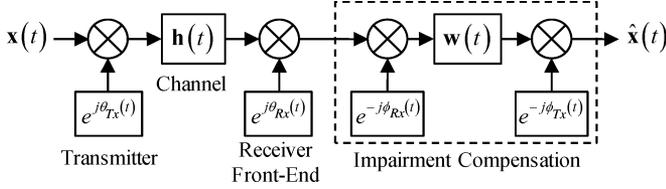


Fig. 12. Channel model in presence of phase noise and LTI impairments.

the tap adjustment $\Delta \mathbf{W}_m$, averting the latency of the FFT/IFFT. Hence, if system performance is not constrained by nonlinearity, and transmitter side DSP is available, OFDM can be an attractive modulation format.

V. LASER PHASE NOISE COMPENSATION

A. Joint Laser Phase Noise Compensation and Linear Equalization

In the presence of laser phase noise and dispersion, a fiber channel has the model shown in Fig. 12, where the phase noise of the transmitter is denoted as $\theta_{Tx}(t)$, and the phase noise of the receiver's LO laser is $\theta_{Rx}(t)$. The ideal compensator for the phase noises and dispersion is channel inversion shown by the dashed box in Fig. 12. For clarity, each block in the channel inverter is shown as continuous-time functions. In DSP compensation, equivalent operations can be performed provided the sampling rate exceeds Nyquist's criterion.

Since rotation by laser phase noise is noncommutable with dispersion, the order of DSP blocks cannot be changed without loss of performance. Moreover, as $\theta_{Tx}(t)$, $\theta_{Rx}(t)$, and $\mathbf{h}(t)$ are time varying, their compensators $\phi_{Tx}(t)$, $\phi_{Rx}(t)$, and $\mathbf{w}(t)$ must be adaptive. From the previous section, to adapt each block, the receiver needs to find the error between the desired output of the block and its actual output. For the Tx phase compensator, the desired outputs are the transmitted symbols, so \mathbf{d}_k is given by Table IV for various adaptive algorithms. Since the linear equalizer precedes the Tx phase compensator, the desired output is $\mathbf{d}_k e^{-j\phi_{Tx,k}}$, and the error signal is $\boldsymbol{\varepsilon}_k e^{-j\phi_{Tx,k}}$. It can similarly be shown that for any multiblock nonlinear system, the error for each block is obtained by backpropagating² the output error $\boldsymbol{\varepsilon}_k$ through all other blocks downstream from it, as shown in Fig. 13(a).

We observe that the Rx phase noise compensator can be viewed as a baseband digital LO controlled by a PLL, where the feedback delay is equal to the sum of the latencies of all the other DSP blocks, including the linear equalizers in the impairment compensation and error feedback paths. This feedback delay is functionally identical to the physical delay of a hardware PLL. Delay reduces linewidth tolerance. It was shown in [31, eq. (41)] that for large feedback delays τ_Δ , the linewidth requirement scales as $2\pi\Delta\nu\tau_\Delta$. Most of the delay in τ_Δ comes from the group delays of \mathbf{w} and \mathbf{w}^{-1} , as these have to be causal, so the linewidth requirement will scale as

²The principle of backpropagating the error to adapt a nonlinear system has been studied in the field of neural networks [30]. The usage of backpropagation here should not be confused with the nonlinearity compensation technique of "backward propagation" introduced in Section VI.

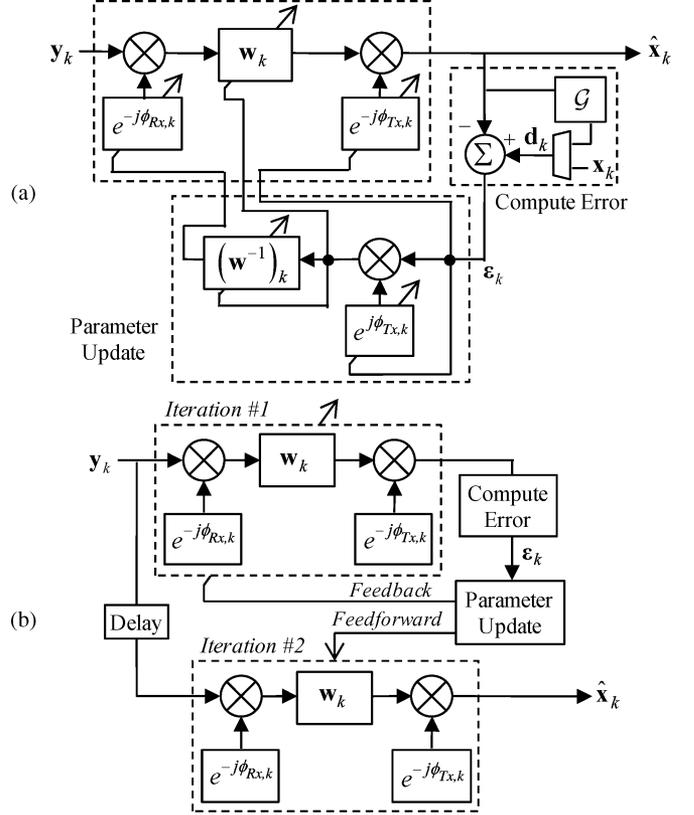


Fig. 13. (a) Adaptive compensation of phase noise and LTI impairments, and (b) iterative compensation using both feedback and feedforward adaptation.

$2\pi\Delta\nu_{LO}N_{cd}T_s$, where N_{cd} is the ISI duration of CD given by (19). This scaling holds for both SC and OFDM systems, as the prefix length of OFDM is proportional to N_{cd} . In contrast, the linewidth requirement for the Tx laser scales as $2\pi\Delta\nu_{Tx}T_s$, as the feedback path does not enclose the filters \mathbf{w} and \mathbf{w}^{-1} . Hence, for a receiver whose parameters are adapted using feedback only, as shown in Fig. 13(a), the linewidth requirement on the LO is much more severe than the Tx. Neglecting nonlinear effects, if Rx phase noise is the limiting factor, 0% residual dispersion at the receiver is favored.

Laser phase noise tolerance can be increased by eliminating feedback. Consider Fig. 13(b), where the error feedback path is relabeled as a "parameter update" circuit. In addition, a copy of the received signal is delayed by the feedback latency. The receiver uses the parameters obtained in the initial adaptation (iteration #1) and processes the received signal a second time (iteration #2) to produce a better estimate of the transmitted signal. This additional signal path represents "feedforward" operation. It is possible to increase the number of iterations of impairment compensation + error estimation + parameter update to obtain better performance.

B. Zero Residual Dispersion

When the channel has zero residual CD, zero differential group delay, and no PDL, the channel's impulse response is at most an instantaneous unitary matrix \mathbf{Q} . The received signal is

$$\mathbf{y}(t) = [\mathbf{Q}\mathbf{x}e^{j\theta_{Tx}(t)} + \mathbf{n}(t)] e^{j\theta_{Rx}(t)}. \quad (34)$$

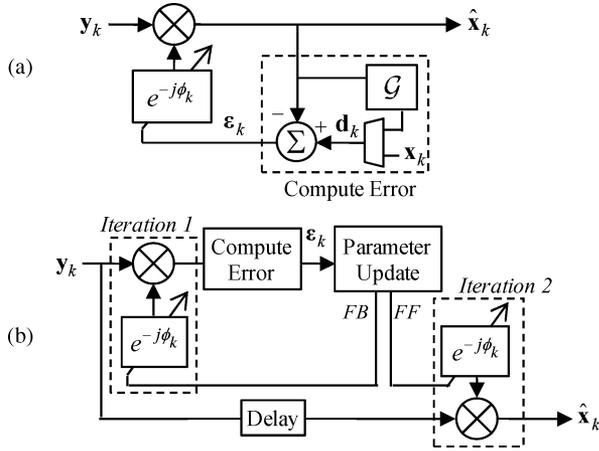


Fig. 14. (a) Digital PLL, and (b) digital feedforward carrier recovery.

Since $\mathbf{n}'(t) = \mathbf{Q}^{-1}\mathbf{n}(t)e^{j\theta_{\text{Rx}}(t)}$ has the same statistics as $\mathbf{n}(t)$, system performance is characterized by $\mathbf{y}'(t) = \mathbf{Q}^{-1}\mathbf{y}(t) = \mathbf{Q}^{-1}\mathbf{x}(t)e^{j(\theta_{\text{Tx}}(t)+\theta_{\text{Rx}}(t))} + \mathbf{n}'(t)$, and the parameter of interest is the beat linewidth normalized to the symbol interval: $\Delta\nu/T_s = (\Delta\nu_{\text{Tx}} + \Delta\nu_{\text{Rx}})T_s$. Thus, Fig. 13(a) reduces to Fig. 14(a), which is the canonical model of a PLL, with the digital Rx phase compensator being a digital LO, which enables the LO laser to be free running. Alternatively, a digital LO can be omitted if the error signal directly controls the hardware optical or electrical LO instead, which results in a traditional PLL where the feedback circuit attempts to make the phase difference $\theta_{\text{Rx}}(t) - \theta_{\text{Tx}}(t)$ as small as possible [31], [32]. PLLs for coherent optical systems was studied for phase-modulated formats in [31] and for arbitrary constellation in [33].

Under the same zero-dispersion assumption, the iterative channel inverter in Fig. 13(b) becomes Fig. 14(b), which is the feedforward carrier recovery (FFCR) structure studied in [34]. Since the equalizers \mathbf{w} and \mathbf{w}^{-1} are not needed, the two phase noise compensators combine into a single phase estimator ϕ_k that tries to match $\theta_{\text{Tx},k} + \theta_{\text{Rx},k}$. FFCR was studied for phase-modulated formats in [35] and [36] and arbitrary constellation in [34]. FFCR has better performance than a PLL because ϕ_k can be optimized using both past and future received symbols, whereas in the PLL, the LO phase only depend on past symbols, as it uses only feedback.

Like the adaptive equalizer that mitigates LTI impairments, the FFCR circuit for mitigating laser phase noise in Fig. 14(b) also has error computation and parameter update blocks that serve identical functions to their counterparts in the adaptive equalizer. Hence, FFCR can use the same techniques of DA, DD, and blind algorithms to adapt the phase noise compensator. The first two techniques are identical to Section IV-C, where DA adaptation uses training symbols to compute the error signal, while DD adaptation uses output decisions. Blind adaptation is qualitatively different. Previously, the Godard algorithm was concerned with the radius of the output signal, and was phase independent. In FFCR, on the other hand, the output phase is of critical interest. For phase-modulated constellations, nondata-aided (NDA) adaptation is commonly used. The Viterbi–Viterbi algorithm exploits the rotational symmetry of M -ary PSK to estimate phase by raising the signal to the M th power and taking

its argument [37]. This avoids the need for detection or training symbols. NDA adaptation has poorer performance compared to DD adaptation [34], just as the CMA has poorer performance compared to DD in adaptive linear equalization.

C. Nonzero Residual Dispersion

In the presence of both dispersion and laser phase noise, the iterative structure of Fig. 13(b) will yield better performance compared to Fig. 13(a), but owing to its complexity, it has not been studied. Various approximations of the iterative structure have been proposed. Most commonly, the Rx phase noise compensator is omitted due to the complexity of backpropagating the output error through \mathbf{w}^{-1} . In the absence of an Rx phase noise compensator, the time-varying phase of the LO laser will “smear” across the linear equalizer coefficients, degrading its performance. This degradation grows with the increasing duration of residual ISI. The power penalty of equalization-enhanced LO phase noise was studied in [38] and [39]. In particular, Shieh and Ho [39] found that whereas Tx phase noise favors the use of high symbol rate to reduce $2\pi\Delta\nu_{\text{Tx}}T_s$, for LO phase noise, where the ISI duration $N_{\text{cd}} \propto 1/T_s^2$, the LO linewidth tolerance $2\pi\Delta\nu_{\text{LO}}N_{\text{cd}}T_s$ scales as $1/T_s$, thus favoring a low symbol rate.

With the removal of $\phi_{\text{Rx}}(t)$, the linear equalizer and Tx phase noise compensator in Fig. 13(a) can be combined into time-varying linear equalizer $\tilde{\mathbf{w}}_k = \mathbf{w}_k e^{j\phi_{\text{Tx},k}}$. If phase noise evolves sufficiently slowly, the adaptive algorithms discussed in Section IV-C will automatically adjust for the evolving phase, eliminating the need for a separate laser phase noise compensator. This configuration was considered in [40]. When residual ISI has long duration, and laser phase noise is significant, this structure is suboptimal because the time variation in $\mathbf{h}(t)$, which arises from PMD and changes on the order of kilohertz, is much slower than laser phase noise which evolves at a rate of tens or hundreds of kilohertz. A simplified structure to Fig. 13(a), where a slower adapting TDE for PMD compensation is followed by a faster adapting FFCR for laser phase noise compensation, was studied in [41].

D. Carrier Recovery for OFDM

For OFDM, the most commonly used technique for laser phase noise mitigation is by insertion of pilot symbols at designated subcarriers, as shown in Fig. 15. It is assumed that the linewidths of the Tx and LO lasers are sufficiently narrow such that over the duration of an OFDM symbol, the changes in $\theta_{\text{Tx}}(t)$ and $\theta_{\text{Rx}}(t)$ are negligible, thus it may be assumed that $\theta = \theta_{\text{Tx}}(t) + \theta_{\text{Rx}}(t)$ is constant. In the absence of noise and other distortions, $\hat{\theta}$ can be estimated from the phase of the received pilot. The receiver can then derotate the entire OFDM symbol by $\hat{\theta}$. In the presence of noise, performance can be improved by increasing the number of pilots. This technique is called “common phase estimation” (CPE). The system model for CPE is shown in Fig. 16. The receiver consists of an FDE followed by a slow Tx phase noise compensator. LO phase noise compensation is omitted. It is observed that the operations performed do not exactly invert the total response of the transmitter, channel, and receiver front end. Hence, even with no LO phase noise, perfect linear equalization, and negligible prefix length, the blocks between the Tx laser and the Rx phase noise

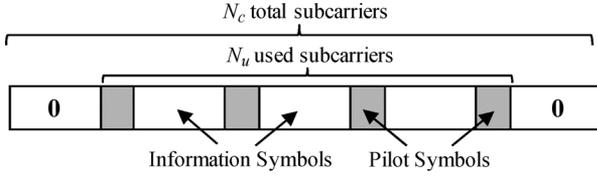


Fig. 15. OFDM carrier recovery using pilot symbols.

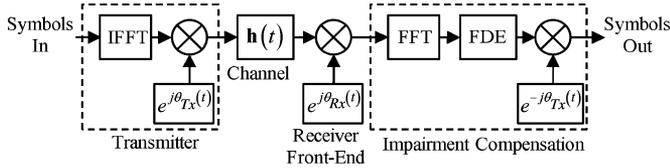


Fig. 16. OFDM carrier recovery using pilot symbols.

TABLE VI
LINWIDTH SCALING FOR DIFFERENT DISPERSION MAPS
AND TRANSMISSION FORMATS

	Single-Carrier (using FFCR)	OFDM (using pilot symbols & CPE)
Zero residual dispersion at Rx	Tx Laser: $\Delta\nu_{Tx} \propto 1/T_s$ LO Laser: $\Delta\nu_{Rx} \propto 1/T_s$	Tx Laser: $\Delta\nu_{Tx} \propto 1/T_{ofdm}$ LO Laser: $\Delta\nu_{Rx} \propto 1/T_{ofdm}$
Nonzero residual dispersion at Rx	Tx Laser: $\Delta\nu_{Tx} = 1/T_s$ LO Laser: $\Delta\nu_{Rx} = 1/(N_{cd}T_s)$	Tx Laser: $\Delta\nu_{Tx} \propto 1/T_{ofdm}$ LO Laser: $\Delta\nu_{Rx} \propto 1/T_{ofdm}$

compensator amount to an N_c -point FFT operation, which is a linear filter with a duration of N_c samples. Therefore, the Tx linewidth tolerance scales as $2\pi\Delta\nu_{Tx}T_{ofdm}$. It can similarly be shown that the LO linewidth tolerance has the same scaling as $2\pi\Delta\nu_{Rx}T_{ofdm}$, since the FDE spreads uncompensated LO phase noise over the same duration.

From the point of view of individual subcarriers, laser phase noise degrades performance because it destroys their mutual orthogonality, causing intercarrier interference (ICI). The impact of ICI and CPE has been studied in [42].

The linewidth tolerance scaling for SC and OFDM systems is summarized in Table VI, assuming the use of FFCR and CPE, respectively. We note that to avoid unnecessarily large guard band, $N_c \gg N_{pre} > N_{cd}$, therefore the linewidth requirement of OFDM is more stringent than those for SC.

VI. NONLINEARITY COMPENSATION

A. Impact on Capacity

In real fiber, signal transmission is distorted by a combination of dispersion and nonlinearity as described by the NLSE in (1). For low launch power and short transmission distances, the channel can be assumed to be LTI, enabling the use of the transmission model shown in Fig. 2, where linear capacity is given by

$$C_{lin} = \log_2(1 + \gamma_s) \quad \text{b/s/Hz per polarization.} \quad (35)$$

For fixed amplifier noise psd, higher launch power is required to increase C_{lin} . However, whereas SNR $\gamma_s = P_{tx}/(N_{pol}N_0)$ is

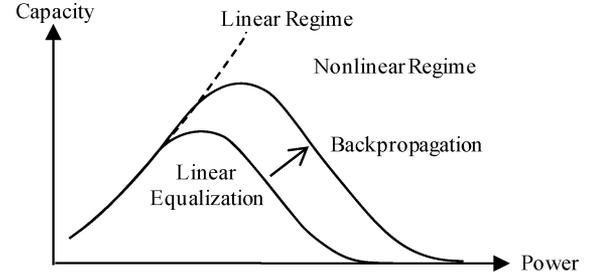


Fig. 17. Channel capacity: linear regime versus nonlinear regime.

only proportional to power, the nonlinear operator of the NLSE is $\hat{N}(\mathbf{E}) \propto |\mathbf{E}|^2$, so the power of nonlinear distortion grows as P_{tx}^2 . Naively, we can consider an “effective SNR” in the nonlinear regime to be $\gamma_s = P_{tx}/(\kappa P_{tx}^2) \propto 1/P_{tx}$, where κ depends on the magnitude of uncompensated nonlinear effects at symbol detection. So in the nonlinear regime, capacity is reduced with further increase in launch power [43], [44]. A characteristic plot of capacity is shown in Fig. 17. Highest capacity occurs at the boundary between the linear and nonlinear regimes. The purpose of nonlinear compensation is to partially mitigate nonlinear effects, reducing κ , thus delaying the onset of the nonlinear regime and increasing the optimum launch power and the achievable capacity.

Traditionally, nonlinear impairments have been categorized into the following types [45]:

- 1) self-phase modulation (SPM): nonlinearity arising in a WDM channel as a result of its own intensity;
- 2) cross-phase modulation (XPM): nonlinearity arising in a WDM channel as a result of the intensity of other channels;
- 3) four-wave mixing (FWM): nonlinearity arising from three interacting fields at different wavelengths producing a nonlinear field at a fourth wavelength;
- 4) nonlinear phase noise (NLPN): nonlinear interaction between signal and noise [46].

The first three types of nonlinear impairments are deterministic given the full field of a WDM signal, while NLPN is stochastic. The strengths of the respective nonlinearities depend strongly on the dispersion map. For example, dispersion causes walkoff between WDM channels, reducing the efficiency of XPM and FWM due to field averaging. Conversely, SPM and XPM can be eliminated using constant amplitude modulation (e.g., PSK or DPSK) in conjunction with negligible fiber dispersion. However, using PSK or DPSK restricts spectral efficiency by encoding information in only one degree of freedom per polarization. We will also see in Section VI-B that low channel dispersion is penalized by NLPN. The capacity of fiber has been studied for various channels and signal processing assumptions in [47]–[50]. The results of greatest theoretical interest are those in [49] and [50], where fiber impairment are compensated using backward propagation (the best method known), and capacity is derived from the mutual information between the transmitter and the receiver assuming the transmitted constellation consists of concentric PSK rings. For transmission over 20×100 -km spans of SMF, Essiambre *et al.* [49] predicted a capacity of 5 b/s/Hz per polarization even for a channel with low residual dispersion per span [49].

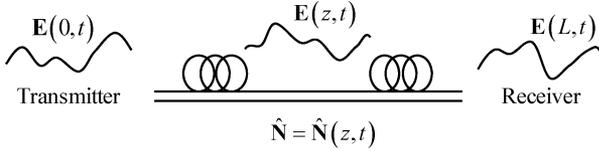


Fig. 18. Nonlinear phase accumulation depends on local signal amplitude.

B. Channel Inversion Using Backward Propagation

In the absence of noise, it can be shown that the NLSE in (1) is an invertible equation. Thus, dispersion and nonlinearity can be simultaneously compensated by solving

$$\frac{\partial \mathbf{E}}{\partial z} = (-\hat{\mathbf{D}} - \xi \hat{\mathbf{N}}) \mathbf{E} \quad (36)$$

where $\hat{\mathbf{D}}$ and $\hat{\mathbf{N}}$ are the linear and nonlinear operators defined in (1). The operation performed by (36) is analogous to passing the received signal $\mathbf{E}(L, t)$ through a fictitious fiber with opposite signs of loss, dispersion, and nonlinearity, yielding an estimate of the transmitted signal $\mathbf{E}(0, t)$. Thus, the technique is commonly referred to as “backward propagation” or “backpropagation” (BP).

In the context of (36), CD and PMD compensation—whether performed optically using DCF or digitally using a linear equalizer—can be viewed as simplified BP taking only into account the linear operator. Similarly, in midspan phase conjugation [51], the second half of the transmission link performs BP on the first half.

BP is an attempt by the receiver to reconstruct the signal’s profile at every point in the fiber in order to undo the correct nonlinear phase shift that was incurred during forward propagation. In the special case of zero-dispersion fiber, the signal maintains its shape throughout transmission except for distortion by amplifier noise. Hence, the received or transmitted signal is a good estimate of the signal profile through the channel, and it is possible to compensate nonlinearity by derotating the received (or transmitted) signal by a phase proportional to its instantaneous amplitude. Nonlinear derotation (NLD) was studied in [52], where it was found that compensating half the nonlinear mean phase shift ($\xi = 0.5$) reduces NLPN variance by 6 dB. In practice, submarine channels that use low-dispersion fibers and near-perfect dispersion compensation after every span to keep accumulated dispersion within a symbol are well modeled as “zero dispersion,” hence NLD can improve performance by about 1 dB [53], [54]. For channels where accumulated dispersion exceeds a symbol period, significant interaction between dispersion and nonlinearity occurs. “Intrachannel four-wave mixing” (IFWM) causes NLD to fail [55] (Fig. 18).

BP was first studied as a transmitter-side electronic precompensation algorithm in [56]–[58] [Fig. 19(a)], since the full complex electric field is always available at the transmitter as the modulator drive signal (Fig. 1). In a coherent system, the electric field is also available at the receiver. Hence, receiver-side electronic BP can also be performed in DSP, and was studied in [55] and [59] [Fig. 19(b)]. The operations performed by the transmitter and the receiver in a coherent system with nonlinear compensation are shown in Fig. 20. Note that NLD is considered

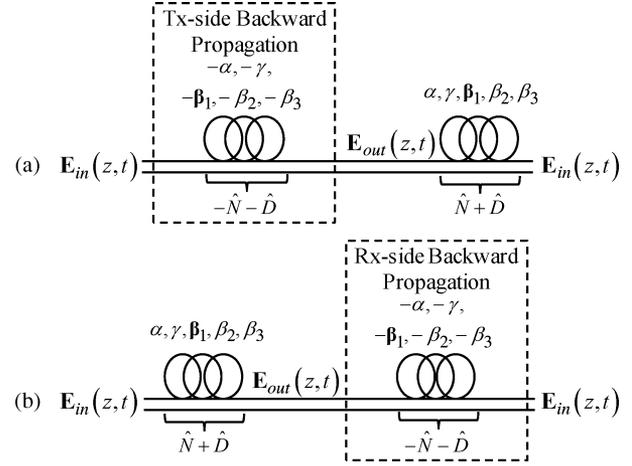


Fig. 19. (a) Transmitter-side backward propagation, and (b) receiver-side backward propagation.

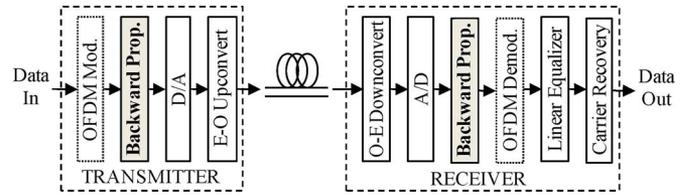


Fig. 20. Nonlinear compensation transmitter and receiver.

as a special case of BP. As BP simultaneously compensates LTI impairments and Kerr nonlinearity, the only other DSP operation required is laser phase noise compensation. In practice, if BP does not take into account all the parameters of the channel, such as unknown PMD in the fiber, a linear equalizer can be inserted following BP to compensate the residual effects.

Since BP (including its NLD approximation) is only concerned with the electric field, it is a universal compensation algorithm independent of the modulation format, which is applicable to both SC and OFDM formats, and can even be used for WDM signals. Simplified versions of Fig. 20 have been employed. For example, transmitter-side NLD was studied in [60], receiver-side NLD was studied in [61], while Lowery [62] showed that performance can be further improved by splitting NLD between the transmitter and the receiver.

In BP, the DSP computes a numerical solution to the NLSE, usually via a split-step Fourier method (SSFM) [63], which divides the fiber channel into sections (Fig. 21). To backpropagate the signal through a section of fiber which extends from $z+h$ to z , we can use the noniterative, asymmetric SSFM (NA-SSFM) [Fig. 22(a)]

$$\mathbf{E}(z, t) \approx \exp\left(-h\xi \hat{\mathbf{N}}(z)\right) \exp(-h\hat{\mathbf{D}}) \mathbf{E}(z+h, t) \quad (37)$$

or we can use a more accurate iterative, symmetric SSFM (IS-SSFM) [Fig. 22(b)]

$$\begin{aligned} \mathbf{E}(z, t) \approx \exp\left(-\frac{h}{2}\hat{\mathbf{D}}\right) \exp\left(-h\xi \frac{\hat{\mathbf{N}}(z+h) + \hat{\mathbf{N}}(z)}{2}\right) \\ \times \exp\left(-\frac{h}{2}\hat{\mathbf{D}}\right) \mathbf{E}(z+h, t). \end{aligned} \quad (38)$$

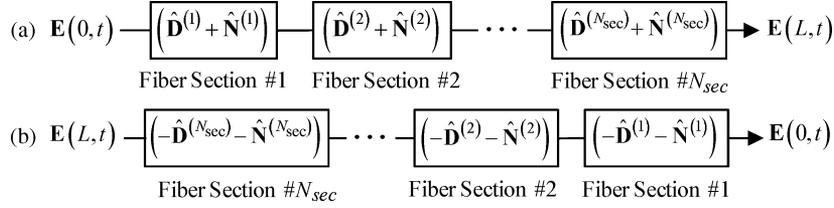


Fig. 21. (a) Forward and (b) backward propagation in a fiber channel modeled as the concatenation of N_{sec} sections.

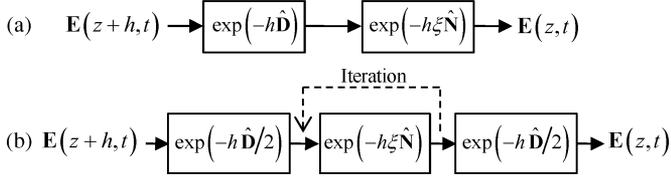


Fig. 22. Backward propagation for one section of fiber using the (a) noniterative asymmetric SSFM, and (b) iterative symmetric SSFM.

The latter algorithm is symmetric because the linear operator is split into two equal parts to be evaluated on either side of the integral of the nonlinear operator over the length of the fiber section. Note in (38) that this integral has been approximated using the trapezoidal rule [63]. An iterative algorithm is required to solve (38), as $\hat{\mathbf{N}}(z)$ depends on $\mathbf{E}(z, t)$, whose value is initially unknown. In the NA-SSFM, the need for iteration is removed by approximating the integral of $\hat{\mathbf{N}}(z)$ with a one-sided rectangular rule. The IS-SSFM is more accurate than the NA-SSFM but is more computationally expensive.

Generally, the linear operator is evaluated in the frequency domain as an all-pass phase filter

$$\begin{aligned} & \mathcal{F} \left\{ \exp(-h\hat{\mathbf{D}})\mathbf{E}(z', t) \right\} \\ &= \exp \left(-h\mathcal{F}\{\hat{\mathbf{D}}\} \right) \mathbf{E}(z', \omega) \\ &= e^{\alpha h/2} e^{-j(\beta_2 \omega^2/2! + \beta_3 \omega^3/3!)h} \mathbf{R}(\theta, \phi) e^{-j\omega(\Delta\tau/2)\sigma_1} \\ & \quad \times \mathbf{R}^H(\theta, \phi) \mathbf{E}(z', \omega) \end{aligned} \quad (39)$$

where $\mathbf{R}(\theta, \phi)$ was defined in (3). The nonlinear operator is most easily evaluated by first transforming into circular polarization coordinates $\mathbf{E}_c(z', t) = \mathbf{\Lambda}\mathbf{E}(z', t)$, where $\mathbf{\Lambda} = (1/\sqrt{2}) \begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}$, and then applying a nonlinear phase shift in the time domain given by

$$\exp(-h\xi\hat{\mathbf{N}}_c)\mathbf{E}_c(z', t) = \begin{bmatrix} e^{-j\xi h \frac{2z}{3}\Phi_1(z', t)} E_{c1}(z', t) \\ e^{-j\xi h \frac{2z}{3}\Phi_2(z', t)} E_{c2}(z', t) \end{bmatrix} \quad (40)$$

$$\Phi_1(z', t) = |E_{c1}(z', t)|^2 + 2|E_{c2}(z', t)|^2 \quad (41)$$

$$\Phi_2(z', t) = 2|E_{c1}(z', t)|^2 + |E_{c2}(z', t)|^2. \quad (42)$$

Finally, we transform back to linear polarization coordinates

$$\exp(-h\xi\hat{\mathbf{N}})\mathbf{E}(z', t) = \mathbf{\Lambda}^H \exp(-h\xi\hat{\mathbf{N}}_c)\mathbf{E}_c(z', t). \quad (43)$$

Solving the SSFM requires using FFTs and IFFTs to switch between frequency and time domain to apply the linear and nonlinear operators, which accounts for the algorithm's high computational cost.

In (36), $0 \leq \xi \leq 1$ is the amount of nonlinearity compensated at each step. Its optimum value depends on system parameters, and is found by numerical simulation. Generally, when dispersion unmanaged pseudolinear transmission is used, and noise is small, the optimum ξ is closer to one, as there is increased confidence that the nonlinear compensation performed at each step undoes the correct phase rotation. Any uncertainty in signal amplitude arising from channel distortions that cannot be measured will degrade performance. These include AWGN, uncompensated PMD, and numerical error.

Since PMD is caused by random fiber birefringence, and is a time-varying effect with a complicated mathematical description, the exact PMD vector through the channel is difficult to estimate. PMD can be suppressed by using better fibers with low PMD. Typical SMF have with PMD as low as 0.05 ps/ $\sqrt{\text{km}}$. Even for a transmission distance as large as 10 000 km, the mean PMD $\overline{\Delta\tau}$ is only 5 ps. Thus, assuming a maximum instantaneous PMD $\Delta\tau_{\text{max}}$ three times the mean value, N_{pmd} remains less than one for 100-Gb/s transmission using polarization-division multiplexed QPSK (PDM-QPSK), so PMD is not a limiting effect—at least for backward propagating a single wavelength channel—and may be mitigated by the post-BP linear equalizer shown in Fig. 20.

Numerical error arises as a result of the digital SSFM solution diverging from the true NLSE solution. This can be caused by using an insufficiently high oversampling, or using a step size that is too large. The sampling rate should satisfy Nyquist's criterion, noting that the signal bandwidth may be higher than expected since nonlinearity creates new frequencies. In [55], it was found that the oversampling rate may need to exceed two to suppress aliasing effects. In the receiver shown in Fig. 20, different sampling rates can be used for the different processing blocks. The step size requirement for a target global accuracy of the SSFM has been studied [64], [65].

Since BP involves applying multiple iterations of the linear and nonlinear operators, its performance is sensitive to any inaccuracies in their representations. For example, the linear operator in (37) is ultimately implemented as an equalizer of finite length (Section IV), where it is critical to minimize the equalizer's amplitude distortion [66].

As an example of the increased nonlinear tolerance obtained with BP, Fig. 23 compares linear equalization with BP for 107-Gb/s SC transmission using PDM-QPSK, and for OFDM, where the channel consists of 25×80 -km spans of SMF. The cases of perfect dispersion compensation, [0% residual dispersion per span (RDPS)], and no dispersion compensation (100% RDPS) are considered. For 0% RDPS using linear equalization only, single-carrier outperforms OFDM because

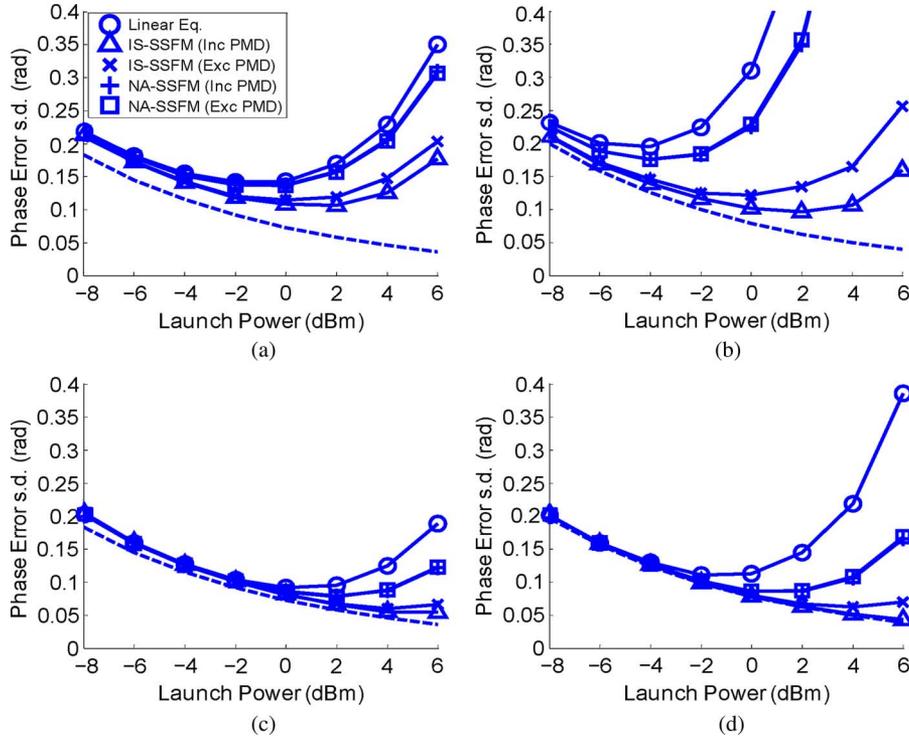


Fig. 23. Performance of BP versus linear equalization for (a) SC and (b) OFDM transmission at 0% RDPS, and (c) SC and (d) OFDM transmission at 100% RDPS.

of the high PAPR of the OFDM. The use of BP reduces their performance difference. More importantly, significantly improved performance is observed with dispersion unmanaged transmission. This is due to: 1) the DCF incurring attenuation and nonlinearity, and 2) residual dispersion enhances walkoff between signal and out-of-band noise, reducing their interaction. The benefit of dispersion-unmanaged transmission has been observed in system simulations in [55] and [67], and experimentally in [29].

For WDM systems, BP can also be used to overcome deterministic cross-channel nonlinearities; in Fig. 23, OFDM is merely a special case of WDM where the subcarriers are orthogonal and closely spaced, so “SPM” in an OFDM signal incorporates all the effects of XPM and FWM between subcarriers. The key limitation to backward propagating a full WDM electric field is that the signal bandwidth may be many terahertz. It is possible to use a bank of phase-synchronized LO lasers for reconstructing the total field of the WDM signal, and then to perform a total field BP. Alternatively, since dispersion renders FWM negligible, BP can be performed individually for each wavelength by solving their reverse NLSEs, taking into account nonlinear coupling via the nonlinear operator. This results in a coupled-field BP approach that facilitates a parallel architecture, and can compensate XPM but not FWM. Both approaches were studied in [59].

VII. CONCLUSION

The ultimate impairment compensation method for a given communications channel is by channel inversion. For optical fiber where signal propagation is impaired by dispersion and Kerr nonlinearity, channel inversion is given by backward prop-

agation of the nonlinear Schrödinger equation. The phase noises of the transmitter and the receiver local oscillator lasers can be mitigated using phase noise compensators inserted before and after backward propagation.

Overcoming time-varying impairments such as polarization-mode dispersion and laser phase noise requires adaptive signal processing. Adaptive algorithms use the error between the desired and the actual output of a function block to train its parameters, with the objective of minimizing a predefined cost function. Depending on whether the desired output is derived from the transmitted symbols, estimates of those symbols, or *a priori* knowledge of the symbols statistics, decision-aided, decision-directed, and blind algorithms can be implemented. For a general multiblock nonlinear system, the error for each block can be found by backpropagating the error at the channel inversion output.

Feedback delay compromises the performance of adaptive algorithms. In particular, equalization-enhanced laser phase noise penalty reduces the linewidth tolerance on the LO laser. Iterative channel inversion employing feedback and feedforward adaptation can lead to improvement in performance.

System design involves making tradeoffs. One such tradeoff is between computational cost and latency. In optical fiber, LTI impairments are most simply described in terms of frequency, so linear equalization has low complexity when performed in the frequency domain. Conversely, Kerr nonlinearity has the simplest description in time; thus, nonlinear phase derotation is best performed in the time domain. Since the symbols of interest in single-carrier transmission form a temporal signal, FDEs have high latency compared to TDEs, due to the need for Fourier transforms.

Another tradeoff is in dispersion map design. Nonlinearity suppression favors using dispersion unmanaged transmission, whereas laser phase noise mitigation favors zero residual dispersion at the receiver.

REFERENCES

- [1] J. M. Kahn, A. H. Gnauck, J. J. Veselka, S. K. Korotky, and B. L. Kasper, "4-Gb/s PSK homodyne transmission system using phase-locked semiconductor lasers," *IEEE Photon. Technol. Lett.*, vol. 2, no. 4, pp. 285–287, Apr. 1990.
- [2] J. M. Kahn, A. M. Porter, and U. Padan, "Heterodyne detection of 310-Mb/s quadriphase-shift keying using fourth-power optical phase-locked loop," *IEEE Photon. Technol. Lett.*, vol. 4, no. 12, pp. 1397–1399, Dec. 1992.
- [3] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurement of a 40-Gb/s coherent system," *Opt. Exp.*, vol. 16, no. 2, pp. 873–879, Jan. 2008.
- [4] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Exp.*, vol. 16, no. 2, pp. 753–791, Jan. 2008.
- [5] S. G. Evangelides, L. F. Mollenauer, J. P. Gordon, and N. S. Bergano, "Polarization multiplexing with solitons," *J. Lightw. Technol.*, vol. 10, no. 1, pp. 28–35, Jan. 1992.
- [6] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proc. Nat. Acad. Sci.*, vol. 97, no. 9, pp. 4541–4550, Apr. 2000.
- [7] J. H. Lee, M. S. Kim, and Y. C. Chung, "Statistical PMD emulator using variable DGD elements," *IEEE Photon. Technol. Lett.*, vol. 15, no. 1, pp. 54–56, Jan. 2003.
- [8] E. Ip and J. M. Kahn, "Power spectra of return-to-zero optical signals," *J. Lightw. Technol.*, vol. 24, no. 3, pp. 1610–1618, Mar. 2006.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 1991.
- [10] J. Armstrong, "OFDM for optical communications," *J. Lightw. Technol.*, vol. 27, no. 3, pp. 189–204, Feb. 2009.
- [11] D. J. F. Barros and J. M. Kahn, "Optimized dispersion compensation using orthogonal frequency-division multiplexing," *J. Lightw. Technol.*, vol. 26, no. 16, pp. 2889–2898, Aug. 2008.
- [12] B. Widrow and S. D. Stearn, *Adaptive Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1985.
- [13] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2033–2043, Aug. 2007.
- [14] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.
- [15] R. D. Gitlin and S. B. Weinstein, "Fractionally spaced equalization: An improved digital transversal equalizer," *Bell Syst. Tech. J.*, vol. 60, no. 2, pp. 275–296, Feb. 1981.
- [16] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice-Hall.
- [17] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [18] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1867–1875, Nov. 1980.
- [19] A. Leven, N. Kaneda, and Y.-K. Chen, "A real-time CMA-based 10 Gb/s polarization demultiplexing coherent receiver implemented in an FPGA," presented at the Opt. Fiber Commun. Conf., San Diego, CA, 2008, paper OTuO2.
- [20] J. Yang, J.-J. Werner, and G. A. Dumont, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 997–1015, Jun. 2002.
- [21] S. Abrar, "Compact constellation algorithm for blind equalization of QAM signals," in *Proc. Int. Netw. Commun. Conf.*, Lahore, Pakistan, 2004, pp. 170–174.
- [22] Y. Chen, T. Le-Ngoc, B. Champagne, and C. Xu, "Recursive least squares constant modulus algorithm for blind adaptive array," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1452–1456, May 2004.
- [23] J. J. Shynk, "Frequency-domain multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.
- [24] M. Kuschnerov, F. N. Hauske, K. Piyawanno, B. Spinnler, A. Napoli, and B. Lankl, "Adaptive chromatic dispersion equalization for non-dispersion managed coherent systems," presented at the Opt. Fiber Commun. Conf., San Diego, CA, 2009, paper OMT1.
- [25] D. Mansour and A. H. Gray, "Unconstrained frequency-domain adaptive filter," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-30, no. 5, pp. 726–734, Oct. 1982.
- [26] J. S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 2, pp. 373–376, Feb. 1990.
- [27] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Exp.*, vol. 16, no. 2, pp. 804–817, Jan. 2008.
- [28] S. J. Savory, G. Gavioli, R. I. Killey, and P. Bayvel, "Transmission of 42.8 Gbit/s polarization multiplexed NRZ-QPSK over 6400 km of standard fiber with no optical dispersion compensation," presented at the Opt. Fiber Commun. Conf., Anaheim, CA, 2007, paper OTuA1.
- [29] G. Charlet, M. Salsi, P. Tran, M. Bertolini, H. Mardoyan, J. Renaudier, O. Bertran-Pardo, and S. Bigo, "72 × 100 Gb/s transmission over transoceanic distance, using large effective area fiber, hybrid Raman-Erbium amplification and coherent detection," presented at the Opt. Fiber Commun. Conf., San Diego, CA, 2009, paper PDPB6.
- [30] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [31] J. R. Barry and J. M. Kahn, "Carrier synchronization for homodyne and heterodyne detection of optical quadriphase-shift keying," *J. Lightw. Technol.*, vol. 10, no. 12, pp. 1939–1951, Dec. 1992.
- [32] S. Norimatsu and K. Iwashita, "Linewidth requirements for optical synchronous detection systems with nonnegligible loop delay time," *J. Lightw. Technol.*, vol. 10, no. 3, pp. 341–349, Mar. 1992.
- [33] E. Ip and J. M. Kahn, "Carrier synchronization for 3- and 4-bit-per-symbol optical transmission," *J. Lightw. Technol.*, vol. 23, no. 12, pp. 4110–4124, Dec. 2005.
- [34] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2675–2692, Sep. 2007.
- [35] S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent demodulation of optical multilevel phase-shift-keying signals using homodyne detection and digital signal processing," *IEEE Photon. Technol. Lett.*, vol. 18, no. 10, pp. 1131–1133, May 2006.
- [36] M. G. Taylor, "Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments," *IEEE Photon. Technol. Lett.*, vol. 16, no. 2, pp. 674–676, Feb. 2004.
- [37] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 4, pp. 543–551, Jul. 1983.
- [38] C. Xie, "Local oscillator phase noise induced penalties in optical coherent detection systems using electronic chromatic dispersion compensation," presented at the Opt. Fiber Commun. Conf., San Diego, CA, 2009, paper OMT4.
- [39] W. Shieh and K.-P. Ho, "Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing," *Opt. Exp.*, vol. 16, no. 20, pp. 15718–15727, Sep. 2008.
- [40] A. Tarighat, R. C. J. Hsu, A. H. Sayed, and B. Jalali, "Digital adaptive phase noise reduction in coherent optical links," *J. Lightw. Technol.*, vol. 24, no. 3, pp. 1269–1276, Mar. 2006.
- [41] I. Fatadin, D. Ives, and S. Savory, "Blind equalization and carrier phase recovery in 16-QAM optical coherent system," *J. Lightw. Technol.*, to be published.
- [42] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: Consequences and solutions," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1988–1996, Nov. 2004.
- [43] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, Jun. 2001.
- [44] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, G. Kramer, and E. C. Burrows, "The capacity of fiber-optic communication systems," presented at the Opt. Fiber Commun. Conf., San Diego, CA, 2008, paper OTuE1.
- [45] G. P. Agrawal, *Fiber-Optic Communication Systems*, 3rd ed. New York: Wiley, 2002.
- [46] J. P. Gordon and L. F. Mollenauer, "Phase noise in photonic communications systems using linear amplifiers," *Opt. Lett.*, vol. 15, no. 23, pp. 1351–1353, Dec. 1990.
- [47] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, Jun. 2001.
- [48] J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for DWDM systems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, no. 2, pp. 259–272, Mar./Apr. 2004.
- [49] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, G. Kramer, and E. C. Burrows, "The capacity of fiber-optic communication systems," presented at the Opt. Fiber Commun. Conf., San Diego, CA, 2009, paper OTuE17.

- [50] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, and G. Kramer, "Exploring capacity limits of fibre-optic communication systems," presented at the Eur. Conf. Opt. Commun., Brussels, Belgium, 2006, paper We1.E.1.
- [51] A. Chowdhury and R.-J. Essiambre, "Optical phase conjugation and pseudolinear transmission," *Opt. Lett.*, vol. 29, no. 10, pp. 1105–1107, May 2004.
- [52] K.-P. Ho and J. M. Kahn, "Electronic compensation technique to mitigate nonlinear phase noise," *J. Lightw. Technol.*, vol. 22, no. 3, pp. 779–783, Mar. 2004.
- [53] K. Kikuchi, M. Fukase, and S. Kim, "Electronic post-compensation for nonlinear phase noise in a 1000-km 20-Gb/s optical QPSK transmission system using the homodyne receiver with digital signal processing," presented at the Opt. Fiber Commun. Conf., Los Angeles, CA, 2007, paper OTuA2.
- [54] G. Charlet, N. Maaref, J. Renaudier, H. Mardoyan, P. Tran, and S. Bigo, "Transmission of 40 Gb/s QPSK with coherent detection over ultralong distance improved by nonlinearity mitigation," presented at the Eur. Conf. Optical Commun., Cannes, France, 2006, paper Th4.3.4.
- [55] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, Oct. 2008.
- [56] K. Roberts, C. Li, L. Strawczynski, M. O'Sullivan, and I. Hardcastle, "Electronic precompensation of optical nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 18, no. 2, pp. 403–405, Jan. 2006.
- [57] R.-J. Essiambre, P. J. Winzer, X. Q. Qang, W. Lee, C. A. White, and E. C. Burrows, "Electronic predistortion and fiber nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 18, no. 17, pp. 1804–1806, Sep. 2006.
- [58] E. Inuzuka, E. Yamazaki, K. Yonenaga, and A. Takada, "Performance of nonlinear interchannel crosstalk pre-compensation at zero-dispersion wavelength using carrier phase-locked WDM," *Electron. Lett.*, vol. 43, no. 13, pp. 729–730, Jun. 2007.
- [59] E. Mateo, L. Zhu, and G. Li, "Impact of XPM and FWM on the digital implementation of impairment compensation for WDM transmission using backward propagation," *Opt. Exp.*, vol. 16, no. 20, pp. 16124–16137, Sep. 2008.
- [60] A. J. Lowery, "Fiber nonlinearity mitigation in optical links that use OFDM for dispersion compensation," *IEEE Photon. Technol. Lett.*, vol. 19, no. 19, pp. 1556–1558, Oct. 2007.
- [61] W. Shieh, X. Yi, Y. Ma, and Y. Tang, "Theoretical and experimental study on PMD-supported transmission using polarization diversity in coherent optical OFDM systems," *Opt. Exp.*, vol. 15, no. 16, pp. 9936–9947, Jul. 2007.
- [62] A. J. Lowery, "Fiber nonlinearity pre- and post-compensation for long-haul optical links using OFDM," *Opt. Exp.*, vol. 15, no. 20, pp. 12965–12970, Sep. 2007.
- [63] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. San Diego, CA: Academic, 2001.
- [64] O. V. Sinkin, R. Holzlohner, J. Zweck, and C. Menyuk, "Optimization of the split-step Fourier method in modeling optical-fiber communication systems," *J. Lightw. Technol.*, vol. 21, no. 1, pp. 61–68, Jan. 2003.
- [65] Q. Zhang and M. I. Hayee, "Symmetrized split-step Fourier scheme to control global simulation accuracy in fiber-optic communication systems," *J. Lightw. Technol.*, vol. 26, no. 2, pp. 302–316, Jan. 2008.
- [66] G. Goldfarb and G. Li, "Efficient backward-propagation using wavelet-based filtering for fiber backward-propagation," *Opt. Exp.*, vol. 17, no. 11, pp. 8815–8821, May 2009.
- [67] X. Liu and R. W. Tkach, "Joint SPM compensation for inline-dispersion-compensated 112-Gb/s PDM-OFDM transmission," presented at the Opt. Fiber Commun. Conf., San Diego, CA, 2009, paper OTuO5.

Ezra M. Ip received the B.E. (honors) degree in electrical and electronics engineering from the University of Canterbury, Christchurch, New Zealand, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 2004 and 2008, respectively.

Currently, he is a member of the research staff at NEC Labs America, Princeton, NJ, where his research interests include coherent optical systems, nonlinear signal processing, and submarine fiber systems. He is the author or coauthor of more than a dozen papers published in international journals and conference proceedings.

Joseph M. Kahn (M'90–SM'98–F'00) received the A.B., M.A. and Ph.D. degrees in physics from University of California at Berkeley, Berkeley, in 1981, 1983, and 1986, respectively.

From 1987 to 1990, he was at AT&T Bell Laboratories, Crawford Hill Laboratory, Holmdel, NJ. He demonstrated multigigabit per seconds coherent optical fiber transmission systems, setting world records for receiver sensitivity. From 1990 to 2003, he was on the faculty of the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, performing research on optical and wireless communications. Since 2003, he has been a Professor of Electrical Engineering at Stanford University, Stanford, CA. His current research interests include single- and multiple-mode optical fiber communications, free-space optical communications, and microelectromechanical systems (MEMS) for optical communications. In 2000, he helped found StrataLight Communications (now Opnext Subsystems), where he served as Chief Scientist from 2000 to 2003.

Prof. Kahn received the National Science Foundation Presidential Young Investigator Award in 1991. From 1993 to 2000, he served as a Technical Editor of the IEEE PERSONAL COMMUNICATIONS MAGAZINE. Since 2009, he has been an Associate Editor of the IEEE/OSA JOURNAL OF OPTICAL COMMUNICATIONS AND NETWORKING.