Digital Coherent Optical Receivers: Algorithms and Subsystems

Seb J. Savory, *Member, IEEE*

(Invited Paper)

***Abstract*—Digital coherent receivers have caused a revolution in the design of optical transmission systems, due to the subsystems and algorithms embedded within such a receiver. After giving a high-level overview of the subsystems, the optical front end, the analog-to-digital converter (ADC) and the digital signal processing (DSP) algorithms, which relax the tolerances on these subsystems are discussed. Attention is then turned to the compensation of transmission impairments, both static and dynamic. The discussion of dynamic-channel equalization, which forms a significant part of the paper, includes a theoretical analysis of the dual-polarization constant modulus algorithm, where the control surfaces several different equalizer algorithms are derived, including the constant modulus, decision-directed, trained, and the radially directed equalizer for both polarization division multiplexed quadriphase shift keyed (PDM-QPSK) and 16 level quadrature amplitude modulation (PDM-16-QAM). Synchronization algorithms employed to recover the timing and carrier phase information are then examined, after which the data may be recovered. The paper concludes with a discussion of the challenges for future coherent optical transmission systems.**

***Index Terms*—Digital communication, polarization.**

I. Introduction

THE SYMBIOTIC combination of digital signal processing (DSP), coherent detection, and spectrally efficient modulation formats has resulted in the digital coherent optical receiver. Due to advances in silicon technology, CMOS analog-to-digital converters, having sampling rates commensurate with current optical line rates have become available [1], allowing ideas devised in the 1990s [2]–[4] to be commercially developed [5]. Not only does the digital coherent receiver offer improved sensitivity, but also it has allowed systems to overcome optical impairments, such as chromatic dispersion or polarization mode dispersion (PMD), which have hitherto hindered the deployment of high-speed systems [6], [7].

While digital coherent receivers have allowed for record transmission results, such as maximum capacity (32 Tb/s) [8], maximum data per wavelength (5.1 Tb/s) [9], and the maximum capacity distance product (112 Pbkm/s) [10], coherent detection also offers benefits at the network level. Not only can digital coherent receivers allow for more robust transmission, and hence, rerouting, but due to the presence of a local oscillator, the receiver is inherently frequency selective permitting new network architectures to be devised [11], [12].

In order to understand the operation of the digital coherent receiver, in this paper, we consider each of the subsystems required. After giving a high-level overview of the subsystems in Section II, we discuss the optical front end, analog-to-digital converter (ADC) and the DSP algorithms necessary to overcome imperfections in the optical front end in Sections III–V, respectively. Sections VI and VII are concerned with compensation of transmission impairments, including the problem of digital polarization tracking and the behavior of the algorithms used. Sections VIII–X discuss the synchronization algorithms employed after which the data may be recovered. The paper concludes with a discussion of the challenges for next generation coherent optical transmissions systems.

II. Subsystems of a Digital Coherent Receiver

In a digital coherent receiver, there are four key subsystems.

1) Optical front end, which linearly maps the optical field into a set of electrical signals.
2) ADC, which converts from the electrical signals into a set of discrete-time-quantized signals at the sampling rate.
3) Digital demodulator, which converts the digital samples into a set of signals at the symbol rate.
4) Outer receiver, which includes error correction and whose functionality is to optimally decode the demodulated signals in order to produce the best estimate of the sequence of bits, which were encoded by the transmitter.

We shall focus on the first three of these subsystems, which form the inner receiver, whose functionality is to produce a "synchronized channel," which is as close as possible to the information theoretic communication channel.

In order to discuss the DSP contained in the digital demodulator, we begin by considering the structural level design of the DSP, including the subsystems listed in Table I.

While for a particular digital coherent receiver, the subsystems employed may differ slightly from those detailed in Table I, they give some indication as to the design choices, which can be made at a structural level, such as the ordering of the

Manuscript received December 23, 2009; revised January 7, 2010; accepted January 8, 2010. Date of publication May 17, 2010; date of current version October 6, 2010. This work was supported by the Joint Regione Piemonte (Italy)–Politecnico di Torino Visiting Professor Scheme, and by the U.K. Engineering and Physical Sciences Research Council.

The author is with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: ssavory@ee.ucl.ac.uk).

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/JSTQE.2010.2044751

TABLE I
FUNCTIONAL SUBSYSTEMS IN A DIGITAL DEMODULATOR

Subsystem	Functionality
De-skew	Aligns the digital signals temporally
Orthogonalization	Maximizes independence between signals
Normalization	Ensures that signals have correct amplitude
Digital equalization	Corrects for channel impairments
Timing recovery	Determines the timing error
Interpolation	Corrects for timing errors
Carrier phase estimation	Compensates for the carrier phase error

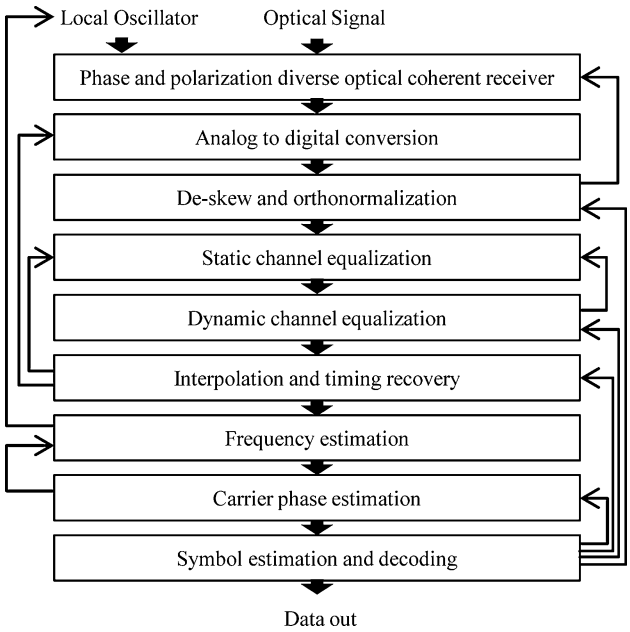

Fig. 1. Subsystems in a digital coherent receiver, including the possible feedback paths.

subsystems.[1] Additional considerations at the structural level include the combining or partitioning of subsystems, for example, the carrier phase estimation may be separated into estimating the frequency offset and the residual carrier phase, whereas in a conventional wireless communication system, it is possible to perform joint carrier phase and symbol synchronization [13].

We shall focus on the structural and algorithmic level rather than the implementation level, where the finite resources, such as machine precision and clock speed are considered. In order to discuss performance of the digital coherent receiver at the algorithmic level, we will restrict our discussion to one possible realization, where the subsystems are arranged, as in Fig. 1, including the "inner receiver" and the "outer receiver," which performs symbol estimation and decoding.

As can be seen in Fig. 1, there are numerous feedback paths between the subsystems. Some of these paths occur naturally, such as between the phase estimation subsystem and frequency

[1]We will focus on a single-carrier system, rather than a multicarrier format, such as orthogonal frequency division multiplexing (OFDM); however, the design methodology is similar.

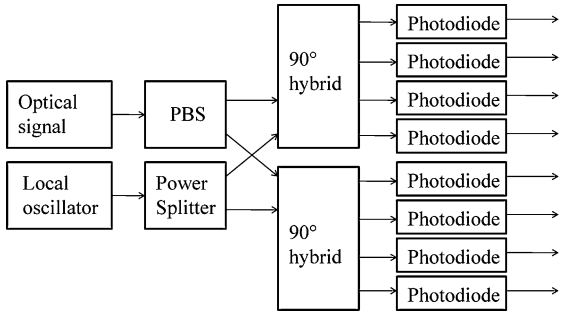

Fig. 2. 2×8 port receiver.

estimation subsystem, however, other paths depend on the algorithms employed. For example, feedback from the symbol estimation and decoding subsystem is required for data-aided algorithms, but not for blind algorithms. Likewise, for synchronous sampling at the baud rate, feedback would be required from the timing-recovery subsystem to the ADC subsystem, which could be omitted for asynchronous sampling. In the subsequent sections of this paper, we shall discuss each of these subsystems independently, forming a basis for understanding the operation of a digital coherent receiver employing feedback between the subsystems.

III. PHASE AND POLARIZATION DIVERSE OPTICAL COHERENT RECEIVER

The functionality of the optical front-end, being both phase and polarization diverse, is to linearly map the incoming optical field into the electrical domain [14]. To realize this functionality, the architecture is shown in Fig. 2, is often employed, which employs a pair of 90° hybrids, one for each component of polarization.[2]

If we consider that the electric field of the incoming optical signal is of the form $[E_x, E_y]^T$ and the local oscillator $E_{\rm lo}$, then following the photodiode, the three complex signals will give rise to a total of nine real quantities, which are[3] $|E_x|^2$,$|E_y|^2$,$|E_{\rm lo}|^2$, $\mathrm{Re}\{E_x E_{\rm lo}^*\}$, $\mathrm{Im}\{E_x E_{\rm lo}^*\}$, $\mathrm{Re}\{E_y E_{\rm lo}^*\}$, $\mathrm{Im}\{E_y E_{\rm lo}^*\}$, $\mathrm{Re}\{E_x E_y^*\}$, and $\mathrm{Im}\{E_x E_y^*\}$. Hence, if we consider these product terms as the input to the subsystem, it is evident that we can model an optically diverse hybrid, by a linear scattering matrix S_h, which may be simplified yet further by noting that for a polarization beam splitter with high extinction ratio, there will be no beating between the two components of the polarization, and hence, $\mathrm{Re}\{E_x E_y^*\}$ and $\mathrm{Im}\{E_x E_y^*\}$ may be neglected. Using the scattering matrix for the 90° hybrid [15], it can be shown that the eight outputs $\mathbf{v} = S_h \mathbf{u}$ of the 2×8 port

[2]In Fig. 2, the power splitter can also be realized using a second polarization beam splitter, assuming the state of polarization of the local oscillator is correctly chosen.

[3]In principle, it might be necessary to consider terms of the form $E_i(t)E_j^*(t-\tau)$, where $i, j \in \{x, y, \mathrm{lo}\}$ and τ is a delay, however, these may be neglected in practice, since we note for practical systems, the phase noise is sufficiently low, such that $E_{\rm lo}(t) \approx E_{\rm lo}(t-\tau)$ and there will be no beating between the two components of the polarization due to the polarization beam splitter (PBS), such that $E_x(t)E_y^*(t) \approx E_x(t)E_y^*(t-\tau) \approx 0$, etc.

receiver are given by[4]

$$\mathbf{v} = \underbrace{\begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} \\ -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} \\ 0 & -\frac{1}{2} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & -\frac{1}{2} & 0 & 0 & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{8} \end{bmatrix}}_{S_h} \underbrace{\begin{bmatrix} \mathrm{Re}\{E_x E_{\mathrm{lo}}^*\} \\ \mathrm{Im}\{E_x E_{\mathrm{lo}}^*\} \\ \mathrm{Re}\{E_y E_{\mathrm{lo}}^*\} \\ \mathrm{Im}\{E_y E_{\mathrm{lo}}^*\} \\ |E_x|^2 \\ |E_y|^2 \\ |E_{\mathrm{lo}}|^2 \end{bmatrix}}_{\mathbf{u}} \tag{1}$$

where $\mathbf{u}$ is the input vector of product terms and S_h is the power scattering matrix of the hybrid. These outputs are then fed into a suitable electrical network, which may also be represented by a scattering matrix. For a receiver based on balanced or single-ended photodiodes, the scattering matrix of the electrical network is given by

$$S_e = \begin{bmatrix} 1 & -b & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -b & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -b & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -b \end{bmatrix} \tag{2}$$

where $b = 0$ for the single-ended case, selecting alternate outputs, and $b = 1$ for the balanced case, selecting the difference between pairs of outputs, and we note that

$$S_e(b=1)\mathbf{v} = S_e(b=1)S_h\mathbf{u} = \begin{bmatrix} \mathrm{Re}\{E_x E_{\mathrm{lo}}^*\} \\ \mathrm{Im}\{E_x E_{\mathrm{lo}}^*\} \\ \mathrm{Re}\{E_y E_{\mathrm{lo}}^*\} \\ \mathrm{Im}\{E_y E_{\mathrm{lo}}^*\} \end{bmatrix} \tag{3}$$

such that all of the direct-detection terms $|E_x|^2$, $|E_y|^2$, and $|E_{\mathrm{lo}}|^2$, are canceled out, with the only remaining terms being the coherently detected products. By rejecting the direct-detection terms, the full potential of the coherent detection to enable a frequency selective receiver is achieved.[5] While this is the case for an ideal 90° hybrid [15], we note there is a matrix H given by

$$H = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix} \tag{4}$$

such that $\mathbf{s} = H\mathbf{v} = HS_h\mathbf{u} = [\mathrm{Re}\{E_x E_{\mathrm{lo}}^*\}, \mathrm{Im}\{E_x E_{\mathrm{lo}}^*\}, \mathrm{Re}\{E_y E_{\mathrm{lo}}^*\}, \mathrm{Im}\{E_y E_{\mathrm{lo}}^*\}, |E_x|^2 + |E_y|^2 + |E_{\mathrm{lo}}|^2]^T$, resulting in five independent quantities, corresponding to the in-phase and quadrature components of the two polarization, in addition to the direct-detection term $|E_x|^2 + |E_y|^2 + |E_{\mathrm{lo}}|^2$. Using this formalism, we can now consider the effect of a nonideal 2×8 port receiver, which can be recast as a multiple input multiple output (MIMO) DSP problem with the objective being to find the matrix $\hat{H}$, such that the signals $\mathbf{s}$ are given by

$$\mathbf{s} = \hat{H}\mathbf{v} = \begin{bmatrix} \mathrm{Re}\{E_x E_{\mathrm{lo}}^*\} \\ \mathrm{Im}\{E_x E_{\mathrm{lo}}^*\} \\ \mathrm{Re}\{E_y E_{\mathrm{lo}}^*\} \\ \mathrm{Im}\{E_y E_{\mathrm{lo}}^*\} \\ |E_x|^2 + |E_y|^2 + |E_{\mathrm{lo}}|^2 \end{bmatrix}. \tag{5}$$

This is an overdetermined MIMO system, and hence, if the fidelity of the digital signal is sufficiently close to the analog waveform, all of these imperfections, can in principle, be compensated digitally, thereby relaxing the requirements on the optical components.[6] While blind source separation techniques, such as independent component analysis could be used [16], it is often possible to simplify this problem, reducing it to finding two pairs of mutually orthogonal components, plus the direct-detection term. Furthermore, in general, we can create a good estimate for $\hat{H}$ based on our *a priori* knowledge of the system,[7] reducing the problem to orthogonalizing pairs of signals, due to the phase in the hybrid not being exactly 90°. In the field of DSP, there are numerous techniques to achieve this for creating orthogonal components, such as the Gram–Schmidt or Löwdin orthogonalization algorithms, which we will discuss in Section V-B.

IV. Analog-to-Digital Conversion

Having mapped the signal from the optical domain into the electrical domain, the next stage is to convert the analog signals into a set of digital signals. From a functional view, we can consider the ADC to be made up of two subsystems, a sampler, which samples the signal in time, converting the continuous-time analog signal into a discrete-time analog signal, followed by a quantizer, which converts the discrete-time analog signal into a finite set of values determined by the bits of resolution in the ADC.[8]

As noted by Schvan *et al.*, in order to implement a high-speed ADC, there are several options [1].

1) *Flash,* where the performance is limited by the clock distribution accuracy and the comparator characteristics.
2) *Flash with track and hold,* which reduces the demand on comparator, with the performance now limited by the track and hold.

[4]In practice, $\mathbf{v} \propto S_h\mathbf{u}$ due to the responsivity of the photodiodes and any subsequent electrical gain, however, we neglect this scale factor for clarity.

[5]The cancelation of the direct-detection terms also allows for a greater dynamic range for the local-oscillator-to-signal ratio and also the cancelation of the relative intensity noise (RIN) on the local oscillator.

[6]At present, however, the fidelity of the digital signal is typically only five effective bits, and hence, may favor the use of precision balanced receivers followed by four ADCs rather than the eight-port network with eight ADCs.

[7]If a calibration procedure was included, this *a priori* knowledge could be increased, thereby, minimizing the difference between a real hybrid and an idealized hybrid.

[8]In an ADC, timing jitter reduces the physical number of bits to an effective number of bits (ENOB).

3) *Time interleaved,* which uses lower speed ADCs, but requires offset, delay, and gain correction.

Of these structures, it is the time-interleaved structure, which lends itself toward CMOS implementation, and hence, integration with the DSP in a single application-specific integrated circuit (ASIC) [17]. As noted, however, the use of a time-interleaved structure requires that the offset, delay, and gain of the lower speed ADCs be compensated, if spurious tones at the clock rate of lower speed ADCs are to be avoided.[9] Nevertheless, since this is a well understood issue for time-interleaved systems, there exist numerous algorithms to overcome these imperfections, which can either be calibrated out or adaptively compensated [18], [19].

Having established that it is possible to realize a high-speed ADC, we now consider the requirements for such a data converter. For a digital communication system, which transmits symbols at a rate S symbols per second, the minimum sampling rate is S (in hertz). In general, however, for asynchronous sampling a sampling rate of $2S$ Hz, is advantageous, giving rise to two samples per symbol, thereby enabling digital timing recovery.[10] While in practice, there may be a slight difference between clocks of the transmitter and the receiver, the signal may be resampled by interpolating of the digital signal, since as noted by Meyr *et al.* "the digital receiver does not need to have a clock frequency equal to the symbol rate $1/T$ as does the analog counterpart" [13].

V. Deskew and Orthonormalization

The functionality of this subsystem is to compensate for the imperfections in the optical front end, where optical path lengths may not be perfectly matched or the hybrid may not be exactly 90°. Signal deskew compensates for the path length mismatches between the signals, synchronizing the digital signals with respect to one another in time. Following on from this, the orthonormalization process compensates for imbalances in amplitude and nonideal 90° hybrids.

A. Deskew Algorithms

In a laboratory prototype, it is not atypical to observe timing mismatches between the optical paths of tens of symbols, such that the resulting digital signals are not synchronized in time. The timing mismatch between the signal paths may be either measured or estimated by cross correlating the signals, with the aim of the deskew algorithm being the compensate for these delays [21]. Since, in general, the timing delays will not correspond to an integer number of samples, we note any delay τ may be written as $\tau = iT_{\mathrm{ADC}} + \mu T_{\mathrm{ADC}}$, being a basepoint delay of $i = \lfloor \tau / T_{\mathrm{ADC}} \rfloor$ samples, followed by a fractional delay μT_{ADC}. While the basepoint delay is straight forward to implement, being a delayed version of the signal, in order to delay by a fraction of a sample period, interpolation is required, which we discuss further in Section VIII.

[9]While synchronous baud rate sampling offers a significant reduction in the DSP complexity, such as halving the number of taps required to compensate chromatic dispersion, at present asynchronous sampling is the preferred solution, with the current state of the art for a CMOS based ADC is 56 GSa/s, providing 2 Sa/symbol for 28 Gbd PDM-QPSK [20].

[10]While synchronous baud rate sampling offers a significant reduction in the DSP complexity, for example, the number of taps required to compensate chromatic dispersion reduces by a factor of four, at present asynchronous sampling is the preferred solution, with the current state-of-the-art for a CMOS-based ADC is 56 GSa/s, providing 2 Sa/symbol for 28 Gbd PDM-QPSK [20].

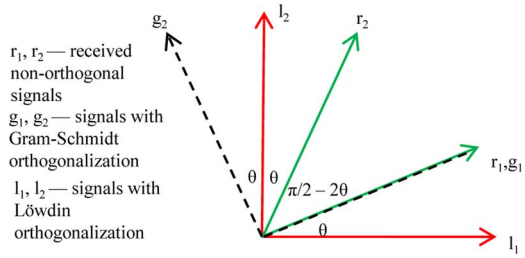

Fig. 3. Visual representation of the Gram–Schmidt and Löwdin orthogonalization algorithms.

B. Orthogonalization Algorithms

In order to recover the original signal using a receiver with a suboptimal hybrid, it is often necessary to use orthogonalization algorithms. In practice, these are applied following digitization, however, in principle, they could be applied via an analog electrical network. To illustrate the effect orthogonalization for two different algorithms, let us consider two transmitted signals t_1 and t_2, which are independent zero mean and of unit variance, such that $E\{t_1 t_2\} = E\{t_1\}E\{t_2\} = 0$ and $E\{t_1^2\} = E\{t_2^2\} = 1$. At the receiver, the hybrid is imperfect, such that the angle between the components is $\pi/2 - 2\theta$, and the received signals r_1 and r_2 are given by $r_1 = t_1 \cos(\theta) + t_2 \sin(\theta)$ and $r_2 = t_2 \cos(\theta) + t_1 \sin(\theta)$. Let us also form the correlation matrix R, which is

$$R = \begin{bmatrix} \langle r_1^2 \rangle & \langle r_1 r_2 \rangle \\ \langle r_2 r_1 \rangle & \langle r_2^2 \rangle \end{bmatrix} = \begin{bmatrix} 1 & \sin(2\theta) \\ \sin(2\theta) & 1 \end{bmatrix}. \quad (6)$$

1) Gram–Schmidt Orthogonalization: The Gram–Schmidt algorithm creates a set of mutually orthogonal vectors, taking the first vector as a reference against which all subsequent vectors are orthogonalized [22]. The outputs $\mathbf{g} = [g_1, g_2]^T$ are determined from the received signals $\mathbf{r} = [r_1, r_2]^T$, by forming the matrix product $G\mathbf{r}$, where G is given by

$$G = \begin{bmatrix} 1 & 0 \\ -\dfrac{\langle r_1 r_2 \rangle}{\langle r_1^2 \rangle} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\sin(2\theta) & 1 \end{bmatrix}. \quad (7)$$

After the signals have been orthogonalized, they are subsequently normalized to unit power, i.e., the second vector is multiplied by $\alpha = \sqrt{\langle r_2^2 \rangle - \langle r_1 r_2 \rangle^2 / \langle r_1^2 \rangle} = \cos(2\theta)$. As can clearly be seen in Fig. 3, the Gram–Schmidt algorithm leaves one vector unchanged, requiring the second vector to be rotated by an angle 2θ.

2) Löwdin Orthogonalization: While the Gram–Schmidt algorithm has been shown to be effective in compensating for imbalance in the 90° hybrid [23], [24], there are other approaches to orthogonalizing the signal. In the Löwdin algorithm, the aim is to create a set of vectors, which are in a least-mean squares sense, closest to the original vectors [25]. It can be shown that

if the correlation matrix is R, the optimal choice of the transformation matrix $L = R^{-1/2}$ [26]. Often this can be achieved via a singular value decomposition of the matrix, however, for the case of the 2×2 signals, considered herein is of the form

$$L = \begin{bmatrix} \dfrac{\cos(\theta)}{\cos(2\theta)} & -\dfrac{\tan(2\theta)}{2\cos(\theta)} \\ -\dfrac{\tan(2\theta)}{2\cos(\theta)} & \dfrac{\cos(\theta)}{\cos(2\theta)} \end{bmatrix} \tag{8}$$

which satisfies the equation $L^2 R = I$ as required. As can be seen in Fig. 3, in contrast to the Gram–Schmidt algorithm, both vectors are equally displaced being rotated by an angle θ.

While the Gram–Schmidt algorithm has been shown to be sufficient for current systems [23], [24], the algorithm increases the impact of quantization noise for the component, which is displaced. In contrast, when symmetric methods are employed, the quantization noise is equally distributed across components. As the spectrally density of the modulation formats increases, alternatives to the conventional Gram–Schmidt procedure, such as the Löwdin [25] or that proposed by Schweinler and Wigner [27] are likely to warrant investigation.

VI. Static-Channel Equalization

One of the key distinguishing features of a digital coherent receiver is its ability to compensate for transmission impairments, in particular, chromatic dispersion and PMD [28], [29]. Much of this is due to the properties of coherent detection, where the optical field is mapped linearly into the digital domain. While in principle equalization could be realized in one subsystem, it is generally beneficial to partition the problem into static and dynamic equalization. Static equalization typically requires large static filters, with dynamic equalization requiring a set of relatively short adaptive filters to compensate for time-varying effects, such as polarization rotations and PMD [7].

A. Linear Compensation

In the absence of nonlinearity, the effect of chromatic dispersion on the pulse $\mathbf{A}(z,t) = [A_x(z,t), A_y(z,t)]^T$ may be modeled by the following differential equation[11] [30]:

$$\frac{\partial \mathbf{A}(z,t)}{\partial z} = \hat{D}\mathbf{A}(z,t) \tag{9}$$

where

$$\hat{D} = j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} \tag{10}$$

with β_2 being the group delay dispersion which is approximately -21 ps^2/km for standard single mode fiber. By solving (9) it can be shown that $\mathbf{A}(z=0,t) = \exp(-\hat{D}L_{\text{total}})\mathbf{A}(z=L_{\text{total}},t)$, and hence the original signal can in principle be recovered from the dispersed signal.[12] By considering the impulse response of the fiber, it can be shown [7], that a signal sampled every T_{ADC} seconds can be recovered by applying a finite impulse response[13] filter to the signal with tap weights $h_{cd}[k]$ given by[14]

$$h_{cd}[k] = \frac{1}{\sqrt{\rho}}\exp\left(-j\frac{\pi}{\rho}\left[k - \frac{N-1}{2}\right]^2\right) \tag{11}$$

where $k \in [0, N-1]$, N is the number of taps given by[15] $N = \lfloor |\rho| \rfloor$ and $\rho = 2\pi\beta_2 L_{\text{total}}/T_{\text{ADC}}^2$. While the dispersion compensating filter could also be realized using a shorter infinite-impulse response filter [31], the FIR realization is often preferred, since it may be implemented efficiently in the frequency domain, e.g., using an overlap and add method [32]. The hardware efficient implementation of the chromatic dispersion compensating filter is an on-going research topic with alternative approaches, such as the subband equalizer structure having been recently explored [33].

B. Nonlinear Compensation

For long distances, the nonlinear transmission may be modeled by the Manakov equation, which is of the form [30]

$$\frac{\partial \mathbf{A}}{\partial z} = (\hat{D} + \hat{N})\mathbf{A} \tag{12}$$

where $\hat{D}$ is given by (10) and $\hat{N}$ is given by

$$\hat{N} = -j\frac{8}{9}\gamma P_0 p(z)\mathbf{A}^{\mathbf{H}}\mathbf{A} \tag{13}$$

where $P_0 p(z)$ is the power profile of the signal along the link and γ is the nonlinear coefficient. By using the Baker–Campbell–Hausdorff formula [34], it can be shown that an approximate solution over a single step of length L, is of the form

$$\mathbf{A}(z=L) \approx \exp(\hat{D}L)\exp\left(\int_0^L \hat{N}dz\right)\mathbf{A}(z=0) \tag{14}$$

which for constant dispersion and polarization independent loss α gives $p(z) = \exp(-\alpha z)$, and $\mathbf{A}^{\mathbf{H}}\mathbf{A} \approx 1$. Hence, if we consider a step size, such that $L\alpha \gg 1$, we find

$$\mathbf{A}(z=0) \approx \exp\left(j\frac{8}{9}\gamma\frac{P_0}{\alpha}\right)\exp(-\hat{D}L)\mathbf{A}(z=L) \tag{15}$$

i.e., to recover the original signal, linear compensation of the chromatic dispersion is applied as per the previous section, followed by an instantaneous nonlinear phase shift [35]. To date, much of the nonlinear compensation is based on (15), with the step-size ranging from the entire length of a system

[11] Herein, we have adopted the electrical engineering convention for Fourier transforms and the associated electric field, such that the field is of the form $E(z,t) = A(z,t)\exp(j[\omega t - \beta z])$ in contrast to the physicists convention that $E(z,t) = A(z,t)\exp(-i[\omega t - \beta z])$. Conversion is easily achieved by substituting $j = -i$ into the nonlinear Schrödinger equation given in [30].

[12] The operator $\exp(\hat{D}z)$ may be defined such that $\exp(\hat{D}z)\mathbf{A} = \mathcal{F}^{-1}\{\exp(\mathcal{F}\{\hat{D}\}z)\mathcal{F}\{\mathbf{A}\}\}$ where $\mathcal{F}$ denotes the Fourier transform.

[13] The FIR filter is known by a number of names, including feed-forward equalizer, transversal tap filter, however the principle is the same, with the output being a linear combination of delayed versions of the input.

[14] This assumes no windowing is applied however in practice these weights could be convolved with the desired pulse shape for matched filtering.

[15] This is an upper bound for N, with an approximate lower bound given by $N_{lb} = \lfloor |2\pi\beta_2 L_{\text{total}}/(T_{\text{sym}}T_{\text{ADC}})| \rfloor$ where T_{sym} is the symbol period.

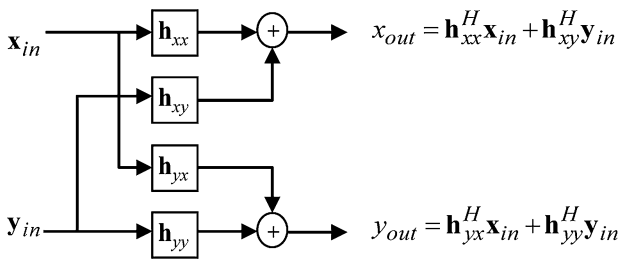

Fig. 4. MIMO equalizer.

to just 2 km [36], with much of the current work employing one step per span.[16] While the complexity may be reduced using wavelets [38], nonlinear compensation based on one step per span may still be prohibitive to implement in hardware. Nevertheless, it provides a benchmark against which simpler compensation schemes may be compared, such as the three block models Hammerstein–Wiener (NLN), the Wiener-Hammerstein (LNL) [39], with the challenge being to improve the nonlinear tolerance of the system without making the DSP prohibitive to implement.

VII. Dynamic-Channel Equalization

The purpose of this subsystem is to compensate for channel impairments, which are time varying, such as the state of polarization or the PMD of the system. The adaptive equalizer required to compensate for these dynamic effects may be realized using a set of four FIR filters [40], as illustrated in Fig. 4.

The functionality of the MIMO filter is to perform the inverse Jones matrix of the dynamic channel,[17] such that the outputs are given by

$$\begin{aligned} x_{\text{out}}[k] &= \mathbf{h}_{xx}^H[k]\mathbf{x}_{\text{in}}[k] + \mathbf{h}_{xy}^H[k]\mathbf{y}_{\text{in}}[k] \\ y_{\text{out}}[k] &= \mathbf{h}_{yx}^H[k]\mathbf{x}_{\text{in}}[k] + \mathbf{h}_{yy}^H[k]\mathbf{y}_{\text{in}}[k] \end{aligned} \tag{16}$$

where for an N tap FIR filter, $\mathbf{h}_{xx}, \mathbf{h}_{yx}, \mathbf{h}_{xy}$, and $\mathbf{h}_{yy}$ are vectors of length N representing the tap weights,[18] with $\mathbf{x}_{\text{in}}$ and $\mathbf{y}_{\text{in}}$ representing a sliding block of N samples to which the filter is applied, such that $\mathbf{x}_{\text{in}}[k] = [x_{\text{in}}[k], x_{\text{in}}[k-1], \ldots x_{\text{in}}[k-N]]$.

In order to understand the salient features of the algorithms employed in dynamic-channel equalization, we reduce the problem to that of digital polarization tracking, which shall be the focus of this section.

A. Signal Model

Let us consider that the modulation format of choice is polarization division multiplexed quadriphase shift keyed (PDM-QPSK) signal, which we may represent mathematically as $\mathbf{E}_{tx}(t) = [E_{tx}^x(t), E_{tx}^y(t)]^T$ being of the form

$$\mathbf{E}_{tx}(t) = \begin{bmatrix} \sum_k (a_k + jb_k) P(t - kT) \\ \sum_k (c_k + jd_k) P(t - kT) \end{bmatrix} \tag{17}$$

where $a_k, b_k, c_k, d_k \in \{-1/\sqrt{2}, 1/\sqrt{2}\}$, and $P(t)$ is the pulse shape. If we normalize the pulse power[19] and consider Nyquist pulse shapes, then synchronous samples at the symbol rate $\mathbf{E}_{tx}[k] = \mathbf{E}_{tx}(t = kT)$ may be written as follows:

$$\mathbf{E}_{tx}[k] = e^{j\vartheta[k]} \begin{bmatrix} 1 \\ e^{jn\pi/2} \end{bmatrix} \tag{18}$$

where $\vartheta[k] = \arg(a_k + jb_k)$ and

$$n = \frac{2}{\pi} \left(\arg(c_k + jd_k) - \arg(a_k + jb_k) \right).$$

One of the key features to note from this representation is that the two QPSK constellations have constant modulus, a feature, which we will exploit in the subsequent discussions.

B. Channel Model

If we assume that the baud rate is sufficiently low that the channel may be considered flat and polarization dependent loss is negligible, then we may model the channel by a unitary 2×2 matrix R, which rotates the horizontal and vertical states of polarization at the transmitter to a pair of arbitrary, but orthogonal states[20]

$$R = \begin{bmatrix} \cos(\theta) & e^{-j\phi} \sin(\theta) \\ -e^{j\phi} \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{19}$$

where 2θ and ϕ are the azimuth and elevation rotation angles, respectively. Using this channel model, the received signal is of the form $\mathbf{E}_{rx}[k] = [E_{rx}^x[k], E_{rx}^y[k]]^T = R\mathbf{E}_{tx}[k]$, which gives

$$\mathbf{E}_{rx}[k] = e^{j\vartheta[k]} \begin{bmatrix} \cos(\theta) + e^{jn\pi/2} e^{-j\phi} \sin(\theta) \\ -e^{j\phi} \sin(\theta) + e^{jn\pi/2} \cos(\theta) \end{bmatrix}. \tag{20}$$

Given our desire to polarization demultiplex the signals, a key question is what choices of θ and ϕ result in the signal being successfully demultiplexed. One possible approach to achieving this is the constant modulus algorithm (CMA), which we shall now discuss.

C. Constant Modulus Algorithm

In the 1980s, the CMA was proposed by Godard [44], as a means of blind equalization for QPSK signals. While the use of the CMA for polarization multiplexed signals was not discussed by Godard, who was concerned with 2-D modulation formats, it has subsequently been extensively been applied to polarization division multiplexed optical communication systems, which are essentially 4-D formats. Nevertheless, with the

[16]Given, in general, only one wavelength channel is received, this results in a nonlinear equalizer, which compensates for intrachannel effects, such as self-phase modulation. However, recent research indicates that the benefit afforded by such an equalizer decreases significantly in a wavelength division multiplexing (WDM) environment [37].

[17]Once the equalizer has converged, the tap weights allow the Jones matrix of the channel to be estimated, hence, giving estimates of the residual dispersion PMD and PDL of the link [41]–[43].

[18]In contrast to the static-channel equalizer, the memory of the equalizer is typically of the order of ten symbols or less.

[19]Later, we shall relax this assumption, however, we retain it here, since it makes the mathematics easier to follow.

[20]In practice, there may also be a phase offset between the two polarization channels. While this will modify the shape of the resulting control surface for data aided algorithms, the salient features, such as the existence of local minima are unchanged by the presence of this differential phase, and therefore, to simplify the presentation, we neglect this effect by setting the differential phase to be zero.

notable exception of Kikuchi [45], there is very little discussion in the literature, as to why this algorithm works. As such herein, we shall aim to demonstrate how this algorithm achieves the functionality of a digital polarization demultiplexing, coming from the perspective of a control surface.

The dual polarization CMA aims to restore the QPSK signal by minimizing the cost functions $\varepsilon_x^2 = (1 - |E_{rx}^x|^2)^2$ and $\varepsilon_y^2 = (1 - |E_{rx}^y|^2)^2$. Therefore, given the form of $\mathbf{E}_{rx}[k]$, we can show that the modulus of the signal is as follows:

$$|\mathbf{E}_{rx}[k]|^2 = \begin{bmatrix} 1 + \sin(2\theta)\cos(\phi - n\pi/2) \\ 1 - \sin(2\theta)\cos(\phi - n\pi/2) \end{bmatrix}. \tag{21}$$

As expected, the total power for the two polarization is unchanged by the unitary rotation, with only the ratio of the powers changing. Using these results, we can, therefore, see that

$$\varepsilon_x^2 = \varepsilon_y^2 = \sin^2(2\theta)\cos^2\left(\phi - \frac{n\pi}{2}\right). \tag{22}$$

Given $\varepsilon_x^2 = \varepsilon_y^2$ in the subsequent, we shall only consider ε_x^2 for which we note there are two cases corresponding to n being odd or even, which gives ε_x^2 as follows.

1) For n even, $\varepsilon_x^2 = \sin^2(2\theta)\cos^2(\phi)$.
2) For n odd, $\varepsilon_x^2 = \sin^2(2\theta)\sin^2(\phi)$.

hence, if we average over sufficient data points, assuming n is equally likely to be odd or even, the resulting control signal is as follows:

$$\left\langle \varepsilon_x^2 \right\rangle = \frac{1}{2}\sin^2(2\theta). \tag{23}$$

This has the desired property that by minimizing the cost function $\left\langle \varepsilon_x^2 \right\rangle$, we separate the two polarizations:

1) $E_{rx}^x \propto E_{tx}^x$ for $\theta = m\pi$;
2) $E_{rx}^x \propto E_{tx}^y$ for $\theta = \pi/2 + m\pi$.

Let us consider that we are able to modify the polarization rotation in the channel by varying θ, which could be achieved in practice by using a polarization controller at the input to the coherent receiver. In this case, a key question is to determine how we should update our initial estimate of θ, to ensure demultiplexing is achieved, in the limit as time tends to infinity. Hence, if $\theta[k]$ is the kth estimate of θ, we wish to determine the value of

$$\theta_\infty = \lim_{k\to\infty} \theta[k]. \tag{24}$$

In order to investigate the convergence, there are several issues related to θ_∞, which require following considerations.

1) Does θ_∞ exist, i.e., does the algorithm converge?
2) Is θ_∞ unique?
3) How sensitive is θ_∞ to the initial conditions?
4) How fast does the algorithm converge to θ_∞ and is this rate of convergence sensitive to the initial conditions?

While there are numerous possible update algorithms [22], much of the current research is focused on stochastic gradient descent algorithms, which we now consider.

1) Stochastic Gradient Descent: One approach is to observe that at the minima

$$\frac{d\left\langle \varepsilon_x^2 \right\rangle}{d\theta} = 0 \tag{25}$$

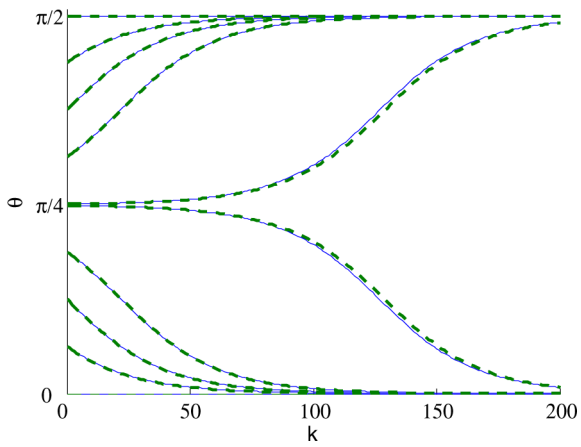

Fig. 5. Evolution of θ determined using a stochastic gradient algorithm with $\mu = 0.01$ with the dashed and thin solid line, and dashed lines denoting solution of the difference (27) and differential (28) equations, respectively.

and close to this minima the function $\left\langle \varepsilon_x^2 \right\rangle$ is convex, and hence, if $\theta[k]$ is an estimate close to the minima a better estimate is obtained by [22]

$$\theta[k+1] = \theta[k] - \mu\frac{d\left\langle \varepsilon_x^2 \right\rangle}{d\theta} \tag{26}$$

where μ determines the rate of convergence of the algorithm. For the case, where $\left\langle \varepsilon_x^2 \right\rangle = 1/2\sin^2(2\theta)$, the update rule gives

$$\theta[k+1] = \theta[k] - \mu\sin(4\theta[k]). \tag{27}$$

By inspection, we note that when $\sin(4\theta[k]) = 0$, $\theta[k+1] = \theta[k]$, and hence, while θ_∞ exists due to the periodic nature of the sine function, it is not unique. Having established the existence of θ_∞, in order to investigate the convergence, we must solve the governing nonlinear difference equation. While we can achieve this numerically, a closed-form analytical solution is possible if we begin by approximating the nonlinear difference equation by a nonlinear differential equation

$$\frac{d\theta[k]}{dk} = -\mu\sin(4\theta[k]) \tag{28}$$

which has the following solution:

$$\theta[k] = \frac{\pi}{2}\mathrm{nint}\left(\frac{2\theta[0]}{\pi}\right) + \frac{1}{4}\arg(f(\theta[0], k, \mu)) \tag{29}$$

where $\mathrm{nint}(x)$ denotes the nearest integer to x and

$$f(\theta[0], k, \mu)) = \sin(4\theta[0]) + j\cosh(4k\mu)\cos(4\theta[0]) + j\sinh(4k\mu))\sin(4\theta[0]). \tag{30}$$

As can be seen in Fig. 5, a bifurcation occurs at $\pi/4$ with two very close initial starting positions leading to very different final solutions, being either $\theta = 0$ or $\pi/2$ in this case. The sensitivity of the final solution to the initial conditions near $\theta = \pi/4$ is typical of nonlinear differential equations, leading to the so-called "butterfly effect," in chaos theory [46]. In addition to this issue, the other key feature to observe is that the time taken to

convergence increases significantly in the region of the bifurcation, where a singularity is observed.

While this may appear a trivial example, it highlights many of the key issues associated with the convergence algorithms, such as the solution of a nonlinear difference equation. Often it is not possible to obtain a closed solution to the governing equations, requiring approximations to be made, or the equations to be linearized about an operating point [47]. While linearization gives an insight as to tracking performance, it hides issues associated with acquisition, such as the presence of multiple minima, leading to one of the most interesting and frustrating features of the blind CMA algorithm when applied to polarization tracking, namely the possibility that the two outputs lock on to the same source [48], [49], being strongly affected by the choice of initial taps weights [50]. It is for these reasons that we pursue a more qualitative assessment of the equalizer, namely the 2-D control surface from which many of the key issues associated with the acquisition can be ascertained.

D. CMA With Arbitrary Initial Tap Weights

At the receiver the incoming signal is of the form

$$\mathbf{E}_{rx}[k] = R\mathbf{E}_{\text{in}}[k] \tag{31}$$

we will, however, apply an equalizer to this using a set of tap weights h, such that

$$\mathbf{E}_{\text{eq}}[k] = h^H \mathbf{E}_{rx}[k]. \tag{32}$$

We consider that the signal has passed through an unknown polarization rotation R given by

$$R = \begin{bmatrix} \cos(\theta) & \exp(-j\phi)\sin(\theta) \\ -\exp(j\phi)\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{33}$$

with the CMA initialized as h_0 given by

$$h_0 = \begin{bmatrix} \cos(\alpha) & \exp(-j\psi)\sin(\alpha) \\ -\exp(j\psi)\sin(\alpha) & \cos(\alpha) \end{bmatrix} \tag{34}$$

such that the initial control surface $J = \langle \varepsilon_x^2 \rangle$ is as follows:

$$\begin{aligned} J = &-\frac{1}{2}\sin^2(2\alpha)\sin^2(2\theta)\cos^2(\phi-\psi) \\ &-\frac{1}{2}\sin(4\alpha)\cos(2\theta)\sin(2\theta)\cos(\phi-\psi) \\ &+\frac{1}{2}(1-\cos^2(2\alpha)\cos^2(2\theta)). \end{aligned} \tag{35}$$

As can be seen from Fig. 6, in the region of interest, the control surface exhibits two distinct optima due to the two possible conditions, which result in the two polarization channels being correctly de-multiplexed. These two regions are separated by line defined by the maximum value of J, which by inspection of (35) takes the value of $1/2$ for which

$$\theta = \frac{\pi}{2} - \frac{1}{2}\arg(\Theta) \tag{36}$$

where $\Theta = -\cos(2\alpha) + j\cos(\phi-\psi)\sin(2\alpha)$.

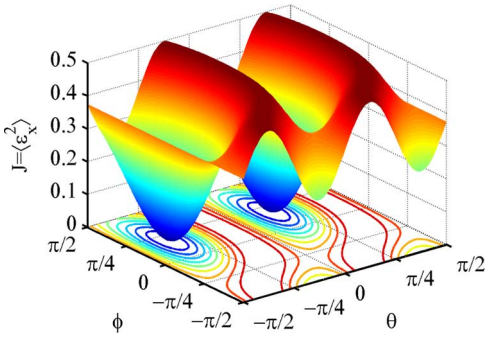

Fig. 6. Control surface J with $\alpha = \pi/6$ and $\psi = \pi/5$.

E. Numerical Investigation

In order to confirm our theoretical insight, we wish to perform a numerical simulation of the system, however prior to this, it is necessary to derive the update algorithm for the CMA with stochastic gradient.

1) CMA Algorithm With Stochastic Gradient: The CMA attempts to minimize $\varepsilon_x^2 = (1-|x_{\text{out}}|^2)^2$ and $\varepsilon_y^2 = (1-|y_{\text{out}}|^2)^2$, often using a stochastic gradient algorithm, such that, for example, the update algorithm for the taps h_{xx} is given by

$$h_{xx} = h_{xx} - \frac{\mu}{2}\frac{\partial}{\partial h_{xx}^*}\varepsilon_x^2 = h_{xx} + \mu\varepsilon_x \frac{\partial}{\partial h_{xx}^*}x_{\text{out}}^* x_{\text{out}} \tag{37}$$

where $x_{\text{out}} = h_{xx}^* x_{\text{in}} + h_{xy}^* y_{\text{in}}$ and μ is the convergence parameter. We define the complex conjugate derivative as follows [22]:

$$\frac{\partial}{\partial h^*} = \frac{1}{2}\left(\frac{\partial}{\partial \text{Re}\{h\}} + j\frac{\partial}{\partial \text{Im}\{h\}}\right) \tag{38}$$

which gives the following update algorithm for the taps:

$$\begin{aligned} h_{xx} &= h_{xx} + \mu\varepsilon_x x_{\text{in}} x_{\text{out}}^* \\ h_{xy} &= h_{xy} + \mu\varepsilon_x y_{\text{in}} x_{\text{out}}^* \\ h_{yx} &= h_{yx} + \mu\varepsilon_y x_{\text{in}} y_{\text{out}}^* \\ h_{yy} &= h_{yy} + \mu\varepsilon_y y_{\text{in}} y_{\text{out}}^*. \end{aligned} \tag{39}$$

2) Simulation Model: To investigate the occurrences of singularities numerically, we consider a PDM-QPSK system with an optical to SNR (OSNR) commensurate with a bit error rate (BER) of 1×10^{-3}. For the case of 100 Gigabit Ethernet (GbE) system, this corresponds to 28 GBd PDM-QPSK signals with an OSNR of 13.3 dB, measured using a noise bandwidth of 0.1 nm.

A set of four bit sequences, each of length 2^{12} bits were used to create a stream of 2^{12} PDM-QPSK symbols.[21] Both θ and ϕ were swept from $-\pi/2$ to $\pi/2$ in steps of $\pi/128$, giving rise to a total of 16 641 simulation points. A convergence parameter of $\mu = 10^{-3}$ was used, cycling through the pattern a total of four times; the first three times being used to ensure the tap weights

[21]In the case of analyzing, the singularity sequences were based on rounding random variables created using the MATLAB rand function, however, we also explored using a 2^{12} de Bruin sequence with a delay of 2^{10} symbols between the four tributaries, which gave the same results.

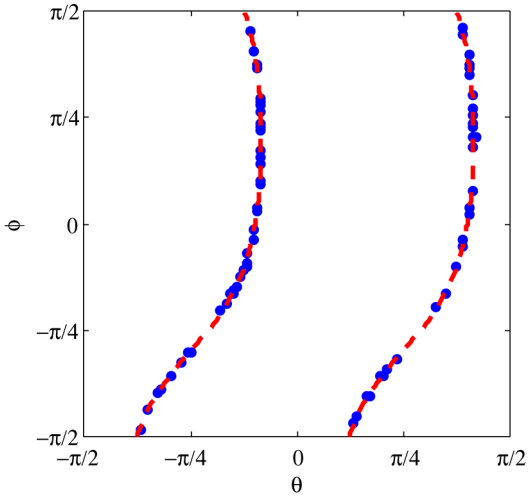

Fig. 7. Analytical prediction (dashed lines) and simulation points (dots), where singularity were observed with $\alpha = \pi/6$ and $\psi = \pi/5$.

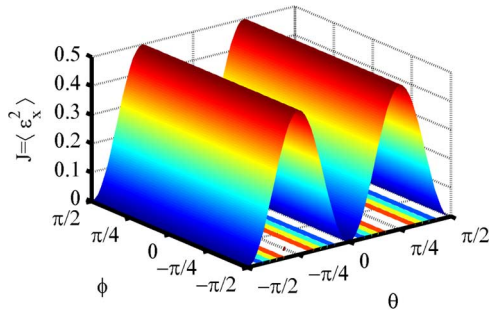

Fig. 8. Control surface for PDM-QPSK with a CMA-based equalizer.

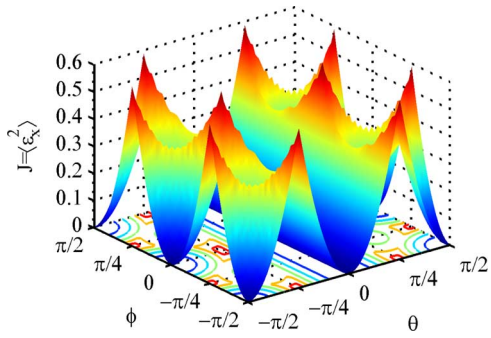

Fig. 9. Control surface for PDM-QPSK with a DD-EQ.

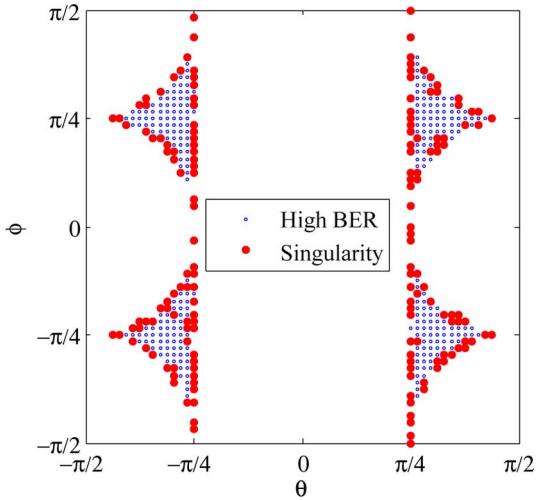

Fig. 10. Location of points with high BER ($>10^{-2}$) c.f. BER $= 10^{-3}$ target and points, where a singularity occurs.

had fully converged, with the final pass being used to measure BER and assess whether a singularity had occurred.

3) Agreement Between Simulation and Theory: As can be seen in Fig. 7, there is excellent agreement between the analytical predictions of where the singularity should occur and the recorded location of where a singularity occurred. Given this, it is clear that the control surface illustrated in Fig. 6 gives significant insight into the convergence of the algorithm. Given this, we now propose to extend this very visual approach to algorithms, such as the decision-directed equalizer (DD-EQ) or one, which employs training sequences for PDM-QPSK.

F. Control Surfaces for PDM-QPSK

1) Constant Modulus Algorithm: While we have already investigated the CMA in the previous sections, in order to perform a comparison with the other methods, herein we derive the control surface. In all cases, we shall assume that a stochastic gradient method is used to achieve convergence of the taps.

The cost function is as follows:

$$\varepsilon_x^2 = (1 - |E_{\text{eq}}^x|^2)^2 \tag{40}$$

which gives the control surface illustrated in Fig. 8.

As previously discussed, this exhibits multiple minima, however, notwithstanding the issue associated with the singularity, the CMA will ensure that the signal is correctly polarization-demultiplexed. Furthermore at the minima point not only is the error gradient zero, but also the value is zero and hence the steady state error is zero.

2) Decision Directed: In a DD-EQ, the output of the equalizer is feed into a decision circuit, such that the cost function is given by

$$\varepsilon_x^2 = \left| \frac{\text{csgn}(E_{\text{eq}}^x)}{\sqrt{2}} - E_{\text{eq}}^x \right|^2 \tag{41}$$

where $\text{csgn}(x)$ is given by

$$\text{csgn}(x) = \begin{cases} 1+j & [\text{Re}(x) > 0, \text{Im}(x) > 0] \\ 1-j & [\text{Re}(x) > 0, \text{Im}(x) < 0] \\ -1+j & [\text{Re}(x) < 0, \text{Im}(x) > 0] \\ -1-j & [\text{Re}(x) < 0, \text{Im}(x) < 0]. \end{cases} \tag{42}$$

As can be seen in Fig. 9, the DD-EQ exhibits features not seen in the CMA surface, namely the presence of nonoptimal local minima. In principle, such an equalizer may become stuck in a local minima, and hence, fail to converge to the correct output. Nevertheless, if the equalizer is preconditioned by using a CMA to bring the tap weights close to the desired minima, e.g., $-\pi/4 < \theta < \pi/4$, then the DD-EQ will converge.

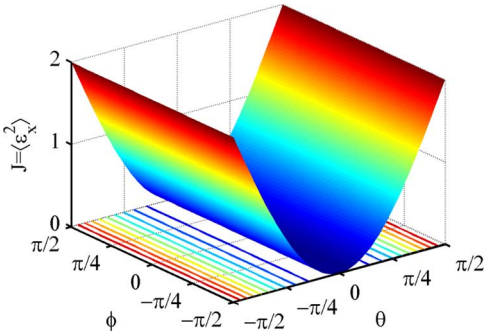

Fig. 11. Control surface for PDM-QPSK with an equalizer employing training sequences.

To investigate further the issue associated with the local minima exhibited by the control surface, we perform a numerical simulation with the same parameters, as those detailed in Section VII-E, other than the angular resolution, which was decreased from $\pi/128$ to $\pi/64$ to reduce simulation time.

As can been seen in Fig. 10, the control surface gives excellent insight into the convergence of the equalizer with the triangular plateaus resulting in a regions of high BER,[22] and the boundaries between regions predicting the location of the points, which give rise to a singularity.

3) Training Based: Thus far the algorithms are essentially blind in that they exploit properties of the modulation format, such as the constant modulus, but have no knowledge of what data was transmitted. An alternative approach is to train the equalizer with a known training sequence such that the equalizer now has complete knowledge of the transmitted data.[23]

In this case, if the original data was x, then equalizer attempts to minimize the function

$$\varepsilon_x^2 = \left| x - E_{\text{eq}}^x \right|^2 . \tag{43}$$

As can be seen from Fig. 11, in contrast to the previously considered cases, the control surface is smooth without local minima in the range considered. Given this control surface, one would expect that the convergence properties would be the best of those considered, albeit at the expense of requiring redundant symbols in order to train the equalizer.[24]

G. Control Surfaces for PDM–16-QAM

Having developed this new insight into the control surfaces for PDM-QPSK, it is possible to extend this approach to other modulation formats, such as PDM–16-QAM. As can be seen in Fig. 12, while 16-QAM does not have constant modulus, it may be considered to be made up of three shells, which is a property that a blind equalizer may exploit [51].

[22]Given the expected BER is 10^{-3}, we define high BER as one, which is greater that 10^{-2}.

[23]We also assume that the linewidth of the laser is sufficiently low that the carrier phase can be considered constant over the set of training symbols.

[24]It should be noted, however, that in current communication systems, there are known framing symbols, which could be exploited to assist the equalizer without reducing the data throughput.

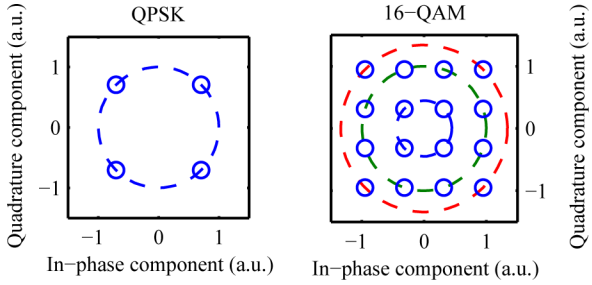

Fig. 12. Constellation diagrams normalized to unit power for QPSK (with unit circle shown) and 16-QAM (with circles of radius $\sqrt{0.2}$, 1 and $\sqrt{1.8}$ shown).

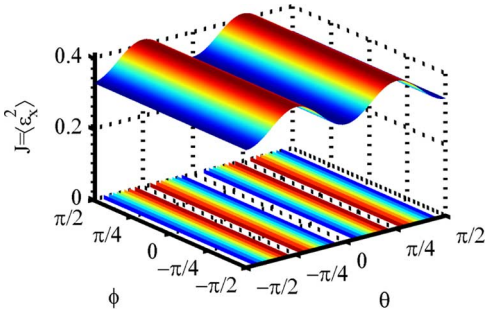

Fig. 13. Control surface for PDM–16-QAM with a CMA-based equalizer.

1) Constant Modulus Algorithm: While a 16-QAM signal is clearly not a constant modulus format, the CMA may be used for preconvergence. As can be seen from Fig. 13, while there is a local minima at the desired location, the error term is nonzero at the optimum with $\min\{\langle \varepsilon_x^2 \rangle\} \approx 0.33$ in this case. Since, the error term does not tend to zero at the optimum, if a stochastic gradient method is used, the equalizer continues to adapt erroneously, introducing noise into the signal, and hence, degrading the performance. As such it is desirable to modify the control surface to ensure that the error signal tends to zero at the optimum point, resulting in the radially directed equalizer (RDE).

2) Radially Directed Equalizer: In order to overcome the limitations associated with the CMA [52], experimental investigations by Winzer *et al.* [53], [54] used an RDE [55], [56], which has an error signal given by

$$\varepsilon_x^2 = (R_0^2 - r^2)^2 \tag{44}$$

where $r = |E_{\text{eq}}^x|$ and R_0 is given by [56]

if $r^2 < 0.6$ **then**
  point belongs to group 1 and $R_0^2 = 0.2$
else
  if $r^2 > 1.4$ **then**
    point belongs to group 3 and $R_0^2 = 1.8$
  else
    point belongs to group 2 and $R_0^2 = 1.0$
  end if
end if

with the signal normalized to unit power on each polarization.

As can be seen from Fig. 14, while minimum value of the control surface is zero, it is well behaved only over a small region, for

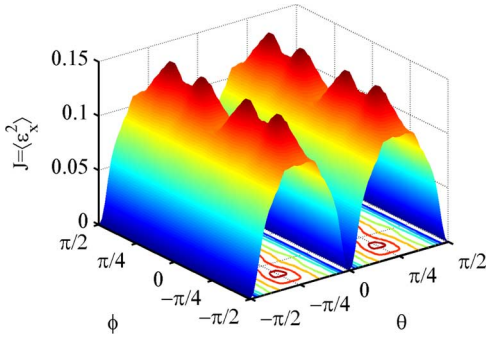

Fig. 14. Control surface for PDM–16-QAM with an RDE.

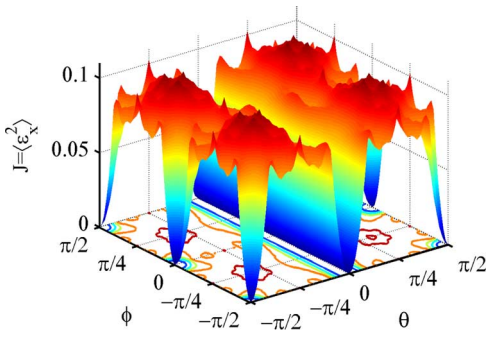

Fig. 15. Control surface for PDM–16-QAM with a DD-EQ.

example, the surface has multiple saddle points for $\theta = \pm\pi/4$. To overcome this limitation, Winzer *et al.* [54] suggested the use of the CMA to bring the equalizer toward the region of convergence, and once the CMA has converged, switching to the RDE; since at the local minima, for the RDE unlike the CMA, the error term is zero. While this is not immediately intuitive from the mathematics, the examination of the control surfaces gives insight as to why this is a sensible scheme.

3) Decision Directed: As for the case of PDM-QPSK, it is also possible to implement a DD-EQ, which has the control surface illustrated in Fig. 15. While, in principle, this algorithm has the minimum tracking error in the presence of noise, the region of convergence is the smallest, such that either CMA or RDE are required to precondition the equalizer.

4) Training Sequences: Hitherto we have considered blind algorithms, however, if a training sequence is available, this may be used to train the equalizer. As can be seen in Fig. 16, the use of training sequences creates a smooth control surface with a global minima at the desired location, being identical in form to Fig. 11, obtained for PDM-QPSK data.

H. Frequency-Dependent Dynamic Channel

Thus far, we have assumed that the dynamic channel is essentially flat, such that it may be reduced to one of polarization tracking. Nevertheless, PMD, for example, results in a frequency-dependent dynamic channel, and therefore, in this section, we consider extending the analysis to compensating such impairments. If we consider the MIMO CMA with a stochastic gradient algorithm, which being very similar to the

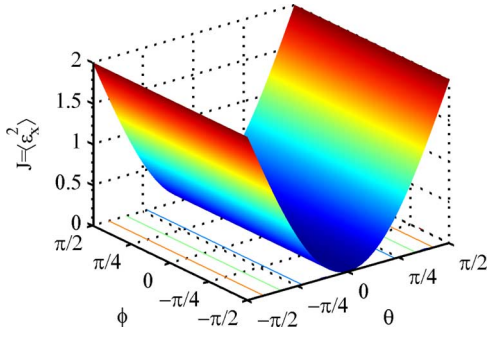

Fig. 16. Control surface for PDM–16-QAM with an equalizer employing training sequences.

single tap equalizer case becomes

$$\begin{aligned}\mathbf{h}_{xx} &= \mathbf{h}_{xx} + \mu\varepsilon_x \mathbf{x_{in}} x^*_{\text{out}} \\ \mathbf{h}_{xy} &= \mathbf{h}_{xy} + \mu\varepsilon_x \mathbf{y_{in}} x^*_{\text{out}} \\ \mathbf{h}_{yx} &= \mathbf{h}_{yx} + \mu\varepsilon_y \mathbf{x_{in}} y^*_{\text{out}} \\ \mathbf{h}_{yy} &= \mathbf{h}_{yy} + \mu\varepsilon_y \mathbf{y_{in}} y^*_{\text{out}}\end{aligned} \tag{45}$$

where $\varepsilon_x = 1 - |x_{\text{out}}|^2$ and $\varepsilon_y = 1 - |y_{\text{out}}|^2$. Such equalizer have been used extensively [7], [57]–[59], in order to increase the robustness of systems to PMD with current commercial systems offering the ability to compensate for a mean differential group delay (DGD) of 25 ps for a 43 Gb/s system using 10.7 Gbd PDM-QPSK. While, in principle, the amount of PMD, which can be compensated is unlimited [60], the limitation is often the complexity of the receiver [61].

I. Overcoming the Singularity Problem

Of the algorithms currently employed for dynamic equalization, the two outputs are converged independently. As such, there is a possibility of the two outputs converging to the same source. At present, several methods have been explored to overcome this limitation, including the use of a cross correlation term in the cost function [48], multiuser CMA [62], independent component analysis [49], magnitude-bounded blind source separation [63], or by the sequentially converging the outputs [50]; however, the optimal solution remains an open research problem. As noted by Haykin "Signal processing is at its best when it successfully combines the unique ability of mathematics to generalize with both the insight and prior information gained from the underlying physics of the problem at hand" [64], and therefore, one solution is to consider the general transfer matrix for the channel [65]

$$T(\omega) = \exp(j\phi(\omega))V(\omega)PU(\omega) \tag{46}$$

where $U(\omega)$ and $V(\omega)$ are unitary matrices, P is the matrix, which represents a partial polarizer (i.e., PDL), and $\phi(\omega)$ is the common phase, which may be a function of ω, due to residual chromatic dispersion. The form of $T(\omega)$ suggests the equalizer structure illustrated in Fig. 17, where $\mathbf{p}_x$ and $\mathbf{p}_y$ compensate for the residual chromatic dispersion, and polarization dependent loss with conjugate symmetry imposed on $\mathbf{u}$ and $\mathbf{v}$ to ensure

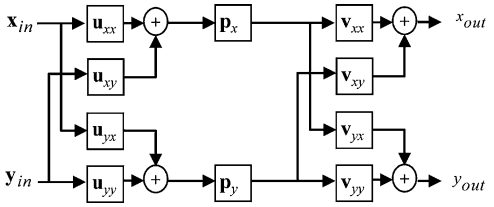

Fig. 17. Modified MIMO equalizer.

they model a unitary matrix, i.e., if $\mathbf{u}_{xx} = (u_{xx}[1], u_{xx}[2], \ldots u_{xx}[N])$, then $\mathbf{u}_{yy} = (u^*_{xx}[N], u^*_{xx}[N-1], \ldots u^*_{xx}[1])$, and $\mathbf{u}_{xy} = (u_{xy}[1], u_{xy}[2], \ldots u_{xy}[N])$, then $\mathbf{u}_{yx} = (-u^*_{xy}[N], -u^*_{xy}[N-1], \ldots - u^*_{xy}[1])$. While the structure of the equalizer in Fig. 17 appears significantly more complex than the original MIMO equalizer of Fig. 4, functionally they are identical, albeit with constraints applied to the tap weights. In principle, it is possible to update all of the tap weights simultaneously resulting in an adaptive cascaded filter. However, the discussion of such filters is outside of the scope of this paper.

VIII. Interpolation and Timing Recovery

Having equalized for the channel impairments, it becomes possible to compensate for the difference between the symbol clock and the ADC sampling rate. As previously mentioned, in principle, it is entirely possible to operate the entire ASIC using only the ADC sampling rate, however, an alternative is to consider partitioning the ASIC, such that one part operates using the ADC clock and the second part uses a derived symbol clock. The task of the interpolating subsystem is to obtain samples $y[k]$ at time $t = t_0 + kT_{\rm sym}$ given samples $x[i]$ at $t = iT_{\rm ADC}$.

A. Interpolation

Interpolation draws on the mathematical theory of approximation theory [66], using Lagrange polynomials or splines to interpolate a sampled function [67]. To illustrate the principle, if $x[i]$ are our samples at $t = iT_{\rm ADC} + \mu T_{\rm ADC}$, where i is an integer and $0 \leq \mu < 1$, then it is possible to create a continuous-time approximation of the form

$$y(t) = \sum_i x[i]\varphi_i(t - [iT_{\rm ADC} + \mu T_{\rm ADC}]) \tag{47}$$

where $\varphi_i(t)$ are a set of interpolation functions.[25] Hence, if we resample this signal at a time $t = kT_{\rm sym} + \varepsilon T_{\rm sym}$, where k is an integer and $0 \leq \varepsilon < 1$, then the samples $y[k]$ will be given by

$$y[k] = \sum_i x[i]\varphi_i(kT_{\rm sym} + \varepsilon T_{\rm sym} - [iT_{\rm ADC} + \mu T_{\rm ADC}]) \tag{48}$$

such that the output is a linear combination of the inputs, and hence, may be realized as an FIR filter.

For the case of linear interpolation, the sample $y[k]$ will be a linear combination of $x[i]$ and $x[i+1]$, such that

$$y[k] = y(m_k, \mu_k) = x[m_k] + \mu_k \left(x[m_k + 1] - x[m_k]\right) \tag{49}$$

where μ_k is the fractional delay, m_k is the basepoint, given by $m_k = \lfloor kw_k + \mu_k \rfloor$, and w_k is our estimate of $T_{\rm sym}/T_{\rm ADC}$. As the sample point is incremented, the revised estimates of basepoint m_{k+1} and the fractional delay μ_{k+1} are given by

$$\begin{aligned} m_{k+1} &= m_k + \lfloor \mu_k + w_k \rfloor \\ \mu_{k+1} &= [\mu_k + w_k]_{\rm mod1}. \end{aligned} \tag{50}$$

Linear interpolation forms the basis of higher order approximations, such as cubic interpolation or piecewise parabolic interpolation [68], [69]. Nevertheless, as the degree of the approximation is increased, it does not necessarily follow that the quality of the approximation improves with a study by Chang *et al.* [70], indicating that piecewise parabolic interpolation outperforms linear or cubic interpolation.

B. Timing Recovery

Digital timing recovery has been a topic of extensive research [13] with both non-data-aided [71] and data-aided [72] algorithms being employed. In order to simplify the interaction between the subsystems, we shall restrict our consideration to nondata-aided methods. One such algorithm aims to maximize the squared modulus $|y(m_k, \mu_k)|^2$ of the interpolated signal $y[k] = y(m_k, \mu_k)$ [13], [70]. Differentiating $|y(m_k, \mu_k)|^2$ with respect to time gives the error signal as follows:

$$\begin{aligned} e(m_k) &= \frac{d|y(m_k, \mu_k)|^2}{dt} \\ &\approx 2\mathrm{Re}\left\{y(m_k, \mu_k)\frac{(y(m_k + 1, \mu_k) - y(m_k - 1, \mu_k))}{2T_{\rm sym}}\right\} \end{aligned} \tag{51}$$

with this error signal going to zero when the signal is synchronized. We may use this error signal $e(m_k)$ to update w_k, our estimate of $T_{\rm sym}/T_{\rm ADC}$, such that

$$w_{k+1} = w_k + \sum_{i=0}^{N-1} c[i]e(m_{k-i}) \tag{52}$$

where the error signal is filtered by an FIR filter of length N with coefficients $c[i]$, which incorporate the convergence factor for this stochastic gradient method.[26]

IX. Frequency Estimation

Given the structural design of the DSP, we have elected to separate the bulk frequency estimation from the carrier recovery. This not only reduces the amount of phase, which the carrier recovery subsystem has to track, but also it improves the efficacy of the carrier recovery, since many phase estimation schemes

[25] Over an infinite-time window, these take the form of $\mathrm{sinc}(x) = \sin(x)/x$ as per the celebrated Shannon–Nyquist sampling theorem, however, for a finite-time window may be realized as Lagrange polynomials or B-splines.

[26] One common approach used by Chang *et al.* [70] is to implement a proportional–integral controller, such that $c[0] = K_i + K_p$ with all other coefficients being K_i.

are only unbiased in the presence of zero frequency offset. If the input signal is of the form

$$x_{\rm in}[k] = x_{\rm sym}[k]\exp(j[\phi[k] + 2\pi\Delta f k T_{\rm sym}]) \tag{53}$$

the task of this subsystem is to estimate Δf.

A. Differential Phase-Based Methods

Given the dominance of PDM-QPSK as a modulation format, we shall first begin with discussing algorithms suitable for this modulation format, prior to discussing higher order modulation formats. For QPSK data, we note that

$$(x_{\rm in}[k]x_{\rm in}^*[k-1])^4 \propto \exp(4j\Delta\phi[k]). \tag{54}$$

In the absence additive noise, $4\Delta\phi$ has a circular Gaussian distribution,[27] due to the laser phase noise with mean $8\pi\Delta f T_{\rm sym}$, such that the probability density function (pdf) is of the form

$$f(4\Delta\phi) = \frac{\exp(\kappa\cos(4\Delta\phi - 8\pi\Delta f T_{\rm sym}))}{2\pi I_0(\kappa)} \tag{55}$$

where κ is related to the linewidth of the laser. Using the pdf, we can estimate the parameter of interest using the maximum likelihood technique [73], which gives an estimate of Δf as follows:

$$\Delta f = \frac{1}{8\pi T_{\rm sym}}\arg\left\{\sum_{k=1}^{N}(x_{\rm in}[k]x_{\rm in}^*[k-1])^4\right\} \tag{56}$$

which is the form of the estimator proposed by Leven *et al.* [74]. Alternatively, another estimator may be obtained by reversing the order of the operations, and iteratively, estimating the frequency offset [75]

$$\Delta f[k] = \Delta f[k-1](1-\mu) + \mu\frac{\arg\left\{(x_{\rm in}[k]x_{\rm in}^*[k-1])^4\right\}}{8\pi T_{\rm sym}} \tag{57}$$

where μ is a convergence factor and $\Delta f[k]$ is the kth estimate of the frequency estimate with the advantage of this method being that it may be implemented using a hardware-efficient realization [75].

B. Spectral Methods

For QPSK data, Δf may be estimated by [76]

$$\Delta f = \arg\max_{\Delta f}\sum_{k=1}^{N} x_{\rm in}^4[k]\exp(-8j\pi k\Delta f T_{\rm sym}) \tag{58}$$

which may be implemented by observing the peak in the spectrum of $x_{\rm in}^4$. In practice, the number of frequency points may be insufficient to give an accurate estimate of the frequency, however, an iterative method may be used to improve this estimate, with this method being equally applicable to QAM formats [77]. While these are essentially feedforward techniques, feedback techniques employing a frequency-controlled loop may also be used, having the advantage that they are agnostic to the modulation format [13], [78].

[27]The circular Gaussian distribution is also known as the Tikhonov distribution and the von Mises distribution.

X. Carrier Recovery

In order to estimate the residual carrier phase, feedforward techniques are often preferable [79], [80]. As with many of the frequency-estimation algorithms, the carrier phase may be estimated using an fourth-order nonlinearity in order to remove the QPSK modulation, giving the estimate of the phase as follows [81]:

$$\phi[k] = \arg\left\{\frac{1}{2N+1}\sum_{n=-N}^{N} w[n]x_{\rm in}^4[k+n]\right\} \tag{59}$$

where $w[n]$ is a weighting function, which depends on the ratio of the additive white Gaussian noise to the laser phase noise [82] (in the absence of laser phase noise $w[n] = 1$, and the estimator becomes that given by Viterbi and Viterbi [81]). The result of the weighting function is to apply a Wiener filter to estimate the phase noise [80], [82], which can approach the performance of an ideal maximum *a priori* (MAP) estimator of the phase. Using these approaches for a 1 dB penalty, a linewidth up to 28 MHz may be tracked for 28 GBd PDM-QPSK [82]. While other estimators such as the Kalman filter have been investigated [83], the efficiency with which the algorithm may be implemented in hardware requires consideration. One particularly hardware-efficient phase estimator is the Barycenter algorithm in which the order of the operations in (59) are reversed to give [84]

$$\phi[k] = \frac{1}{2N+1}\sum_{n=-N}^{N} w[n]\arg\left\{x_{\rm in}^4[k+n]\right\}. \tag{60}$$

As modulation formats move beyond QPSK to QAM, the requirements on the laser linewidth become increasingly stringent [85]. Nevertheless using conventional wireless approaches such as decision-directed phase locked loops, have enabled a linewidth of 1 MHz to be tracked digitally for 14 GBd PDM-16-QAM [52], [54]. Furthermore, for differential 16-QAM and 64-QAM, it has been shown that a digital phase-locked loop can compensate for a residual frequency offset of 1% of the symbol rate [86]. While the digital phase-locked loop presents challenges for CMOS-based parallel implementation, hardware-efficient carrier-recovery schemes have also been proposed with similar performance [87].

In addition to the errors due to the residual phase noise, there is also the possibility of a cycle slip, which can have a catastrophic effect on the performance. As discussed by Taylor [82], in order to reduce the probability of a cycle slip to that of the corrected the BER $= 10^{-18}$, the laser linewidth may need to be reduced by two orders of magnitude, e.g., 600 kHz for 28 GBd PDM-QPSK.[28] In order to minimize the impact of cycle slips, differential decoding may be employed. While this has the effect of increasing the BER by as much as a factor of two, and hence, incurring a modest penalty ($<$1 dB), this often outweighs advantage of relaxed linewidth requirements and simplified symbol decoding.

[28]These stringent linewidth requirements may be optimistic due to the presence of equalization enhanced phase noise [88].

XI. Symbol Estimation and Decoding

Following carrier recovery, the signal may be decoded by the outer receiver. This could take the form of a soft-decision forward error correction (FEC) using a Galois field corresponding to the symbol alphabet, or symbol estimation followed by hard-decision FEC. In current systems, which are based on hard-decision decoding of binary data, symbol estimation and bit decoding is required. For rectangular constellations, such as QAM, this may be achieved by applying a series of decision thresholds to the in-phase and quadrature components separately. While this corresponds to the maximum likelihood symbol estimation for a system limited by additive white Gaussian noise (AWGN), by using nonrectangular decision boundaries, it is possible to improve the performance for systems limited by phase noise, both linear and nonlinear [89].

XII. Challenges for Digital Coherent Transmission Systems

In this final section, we shall consider future challenges for digital coherent transmission systems in which DSP is not only applied at the receiver, but also the transmitter in order to optimize the overall performance.

The first challenge is that of integration of the subsystems, including photonic integration of optical components in the transmitter and receiver allowing the size, cost, and power consumption of the digital coherent modems to be reduced. In addition to photonic integration, digital integration of data converters and associated DSP, is highly desirable. At present, one of the major barriers to the deployment of coherent systems are the resources required to develop an ASIC and while field programable gate arrays (FPGA) offer a means of evaluating algorithms, the lack of integrated data converters make their use prohibitive for commercial systems.[29] The integration of data converters with an ASIC, whose functionality may be controlled via firmware (such as an FPGA) would be a key enabling technology for future digital coherent systems, paving the way for software-defined transceivers. The final level of integration is at the information theoretic level, combining modulation and demodulation with the associated coding and decoding.[30]

The second challenge is that of algorithm design. This includes not only devising algorithms, which can allow the complexity and power consumption of the transceiver to be reduced, employing, for example, baud rate sampling, but also to increase the robustness of systems to imperfections in the photonic components, such as laser phase noise or nonideal hybrids. While the algorithms outlined in this paper are well-suited to current SDH/SONET data payloads, the DSP would need to be fundamentally redesigned to facilitate the detection of bursts of data, and hence, minimizing the loss of data due to the finite-acquisition time. Most of the current research on algorithms draws on methods developed for wireless communication systems (such as OFDM, e.g., [91]–[93]), however, the channel is fundamentally different due to the distributed nature of the nonlinear channel and the presence of non-Gaussian noise [94]. Not only is the optical channel distributed in nature, having both distributed nonlinearities and noise due to the period amplification of the signal, but also the lasers have nonnegligible phase noise. The task of designing CMOS-based photonic digital modems, which allow the information theoretic capacity of the nonlinear dispersive MIMO communication channel with non-Gaussian and signal dependent noise will challenge the community for years to come.

[29] Primarily, due to a data throughput of >1 Tb/s being required for a 100 GbE system (6 bits of resolution $\times$ 4 $\times$ 56 GSa/s), however, the power consumption and cost of this hybrid solution is currently at least an order of magnitude higher than an integrated solution.

[30] One example of this is the turbo equalizer, which combines a MAP receiver with low-density parity codes [90].

XIII. Concluding Remarks

In this paper, we have attempted to outline the subsystems and algorithms, which are required to realize a digital coherent optical receiver. As systems move beyond PDM-QPSK toward higher level modulation formats or to multicarrier techniques, a natural evolution will be for DSP to be employed at both the transmitter and receiver, such that the structural design of the DSP may differ from that presented. Nevertheless, many of the subsystems outlined herein, such as channel equalization and carrier recovery are likely to be present in future photonic digital modems. While the commercialization of photonic digital receivers has begun, it is clear that there remain much research to be done in order to allow a digital coherent optical communication system to achieve the information theoretic nonlinear channel capacity, while offering the functionality that the operators require for future optical networks.

Acknowledgment

The author would like to thank Politecnico di Torino for their support over the summer 2009, when the initial work on the control surfaces was carried out. He would also like to thank Prof. P. Bayvel and Prof. P. Poggiolini for stimulating discussions, and Dr. B. Thomsen, Mr. D. Millar, and Mr. I. Fatadin for their comments on this paper.

References

[1] P. Schvan, J. Bach, C. Fait, P. Flemke, R. Gibbins, Y. Greshishchev, N. Ben-Hamida, D. Pollex, J. Sitch, S.-C. Wang, and J. Wolczanski, "A 24GS/s 6b ADC in 90nm CMOS," in *Proc. Digest Tech. Papers. IEEE Int. Solid-State Circuits Conf., 2008 (ISSCC)*, Feb., pp. 544–634.

[2] F. Derr, "Coherent optical QPSK intradyne system: Concept and digital receiver realization," *J. Lightw. Technol.*, vol. 10, no. 9, pp. 1290–1296, Sep. 1992.

[3] J. Winters, "Equalization in coherent lightwave systems using a fractionally spaced equalizer," *J. Lightw. Technol.*, vol. 8, no. 10, pp. 1487–1491, Oct. 1990.

[4] S. Betti, F. Curti, G. De Marchis, and E. Iannone, "A novel multilevel coherent optical system: 4-quadrature signaling," *J. Lightw. Technol.*, vol. 9, no. 4, pp. 514–523, Apr. 1991.

[5] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Opt. Exp.*, vol. 16, no. 2, pp. 873–879, 2008.

[6] S. J. Savory, G. Gavioli, R. I. Killey, and P. Bayvel, "Electronic compensation of chromatic dispersion using a digital coherent receiver," *Opt. Exp.*, vol. 15, no. 5, pp. 2120–2126, 2007.

[7] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Exp.*, vol. 16, no. 2, pp. 804–817, 2008.

[8] X. Zhou, J. Yu, M.-F. Huang, Y. Shao, T. Wang, P. Magill, M. Cvijetic, L. Nelson, M. Birk, G. Zhang, S. Ten, H. Matthew, and S. Mishra, "32Tb/s

(320×114Gb/s) PDM-RZ-8QAM transmission over 580km of SMF-28 ultra-low-loss fiber," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2009, paper PDPB4.

[9] C. Schmidt-Langhorst, R. Ludwig, D.-D. Gross, L. Molle, M. Seimetz, R. Freund, and C. Schubert, "Generation and coherent time-division demultiplexing of up to 5.1 Tb/s single-channel 8-PSK and 16-QAM signals," in *Proc. Opt. Fiber Commun. Conf*, Mar. 2009, paper PDPC6.

[10] M. Salsi, H. Mardoyan, P. Tran, C. Koebele, G. Charlet, and S. Bigo, "155x100Gbit/s coherent PDM-QPSK transmission over 7,200km," in *Proc. Eur. Conf. Opt. Commun*, Sep. 2009, paper PD2.5.

[11] H. Rohde, S. Smolorz, E. Gottwald, and K. Kloppe, "Next generation optical access: 1 Gbit/s for everyone," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2009, paper 10.5.5.

[12] P. J. Anslow, C. R. S. Fludger, S. Savory, I. Hardcastle, and J. Fells, "Frequency selective coherent receiver for Agile Networks," in *Proc. Eur. Conf. Opt. Commun*, Sep. 2006, paper Mo4.2.4.

[13] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York: Wiley, 1997.

[14] L. Kazovsky, "Phase- and polarization-diversity coherent optical techniques," *J. Lightw. Technol.*, vol. 7, no. 2, pp. 279–292, Feb. 1989.

[15] M. Seimetz and C.-M. Weinert, "Options, feasibility, and availability of 2 × 4 90 degree hybrids for coherent optical systems," *J. Lightw. Technol.*, vol. 24, no. 3, pp. 1317–1322, Mar. 2006.

[16] S. Haykin, *Neural Networks and Learning Machines*. Englewood Cliffs, NJ: Prentice-Hall, 2009.

[17] C. Vogel and H. Johansson, "Time-interleaved analog-to-digital converters: Status and future directions," in *Proc. 2006 IEEE Int. Symp. Circuits Syst., 2006 (ISCAS)*, pp. 3386–3389.

[18] J. Elbornsson, F. Gustafsson, and J.-E. Eklund, "Blind equalization of time errors in a time-interleaved ADC system," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1413–1424, Apr. 2005.

[19] M. Jridi, L. Bossuet, B. Le Gal, and D. Dallet, "New adaptive calibration method for time interleaved analog to digital converters," in *Proc. IEEE Northeast Workshop Circuits Syst., 2007 (NEWCAS)*, Aug., pp. 932–935.

[20] I. Dedic, "56 GS/s ADC: Enabling 100GbE," in *Proc. Opt. Fiber Commun. Conf.*, 2010, p. OThT6.

[21] T. Tanimura, S. Oda, T. Tanaka, T. Hoshida, Z. Tao, and J. Rasmussen, "A simple digital skew compensator for coherent receiver," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2009, paper 7.3.2.

[22] S. Haykin, *Adaptive Filter Theor*. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[23] I. Fatadin, S. J. Savory, and D. Ives, "Compensation of quadrature imbalance in an optical QPSK coherent receiver," *IEEE Photon. Technol. Lett.*, vol. 20, no. 20, pp. 1733–1735, Oct. 15, 2008.

[24] S. H. Chang, H. S. Chung, and K. Kim, "Impact of quadrature imbalance in optical coherent QPSK receiver," *IEEE Photon. Technol. Lett.*, vol. 21, no. 11, pp. 709–711, Jun. 1, 2009.

[25] I. Mayer, "On Lowdin's method of symmetric orthogonalization," *Int. J. Quantum Chem.*, vol. 90, no. 1, pp. 63–65, Oct. 5, 2002.

[26] D. Scofield, "Note on Lowdin orthogonalization and square root of a positive self-adjoint matrix," *Int. J. Quantum Chem.,* vol. 7, no. 3, pp. 561–568, 1973.

[27] H. C. Schweinler and E. P. Wigner. (1970). Orthogonalization methods. *J. Math. Phys.*, [Online]. *11(5)*, pp. 1693–1694. Available: http://link.aip.org/link/?JMP/11/1693/1

[28] 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.

[29] D. Crivelli, H. Carter, and M. Hueda, "Adaptive digital equalization in the presence of chromatic dispersion, PMD, and phase noise in coherent fiber optic systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov. 2003–Dec. 2004, vol. 4, pp. 2545–2551.

[30] G. P. Agrawal, *Lightwave Technology: Telecommunication Systems*. New York: Wiley Interscience, 2005.

[31] G. Goldfarb and G. Li, "Chromatic dispersion compensation using digital IIR filtering with coherent detection," *IEEE Photon. Technol. Lett.*, vol. 19, no. 13, pp. 969–971, Jul. 1, 2007.

[32] M. Kuschnerov, F. Hauske, K. Piyawanno, B. Spinnler, M. Alfiad, A. Napoli, and B. Lankl, "DSP for coherent single-carrier receivers," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3614–3622, Aug. 15, 2009.

[33] K.-P. Ho, "Subband equaliser for chromatic dispersion of optical fibre," *Electron. Lett.*, vol. 45, no. 24, pp. 1224–1226, Nov. 19, 2009.

[34] R. Gilmore, "Baker-Campbell-Hausdorff formulas," *J. Math. Phys.*, vol. 15, no. 12, pp. 2090–2092, 1974.

[35] E. Ip and J. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, Oct. 15, 2008.

[36] F. Yaman and G. Li, "Nonlinear impairment compensation for polarization-division multiplexed WDM transmission using digital backward propagation," *IEEE Photon. J.*, vol. 1, no. 2, pp. 144–152, Aug. 2009.

[37] S. J. Savory, G. Gavioli, E. Torrengo, and P. Poggiolini, "Impact of interchannel nonlinearities on a split-step intrachannel nonlinear equalizer," *IEEE Photon. Technol. Lett.*, vol. 22, no. 10, pp. 673–675, May 2010.

[38] 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, 2009.

[39] D. T. Westwick and R. E. Kearney, *Identification of Nonlinear Physiological Systems*. Piscataway, NJ: IEEE, 2003.

[40] R. Raheli and G. Picchi, "Synchronous and fractionally-spaced blind equalization in dually-polarized digital radio links," in *Proc. IEEE Int. Conf. Commun., 1991 (ICC), Conf. Record.*, Jun., vol. 1, pp. 156–161.

[41] F. Hauske, M. Kuschnerov, B. Spinnler, and B. Lankl, "Optical performance monitoring in digital coherent receivers," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3623–3631, Aug. 15, 2009.

[42] S. Woodward, L. Nelson, M. Feuer, X. Zhou, P. Magill, S. Foo, D. Hanson, H. Sun, M. Moyer, and M. O'Sullivan, "Characterization of real-time PMD and chromatic dispersion monitoring in a high-PMD 46-Gb/s transmission system," *IEEE Photon. Technol. Lett.*, vol. 20, no. 24, pp. 2048–2050, Dec. 15, 2008.

[43] J. Geyer, F. Hauske, C. Fludger, T. Duthel, C. Schulien, M. Kuschnerov, K. Piyawanno, D. van den Borne, E.-D. Schmidt, B. Spinnler, H. de Waardt, B. Lankl, and B. Schmauss, "Channel parameter estimation for polarization diverse coherent receivers," *IEEE Photon. Technol. Lett.*, vol. 20, no. 10, pp. 776–778, May. 15, 2008.

[44] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.

[45] K. Kikuchi, "Polarization-demultiplexing algorithm in the digital coherent receiver," in *Proc. Digest 2008 IEEE/LEOS Summer Topical Meetings*, Jul., pp. 101–102.

[46] L. Smith, *Chaos: A Very Short Introduction*. London, U.K.: Oxford Univ. Press, 2007.

[47] M. Silva and M. Miranda, "Tracking issues of some blind equalization algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 760–763, Sep. 2004.

[48] S. Kun and Z. Xudong, "A SE-CMA based blind equalization for MIMO systems," in *Proc. 7th Int. Conf. Signal Process., 2004 (ICSP)*, Aug./Sep., vol. 2, pp. 1674–1677.

[49] H. Zhang, Z. Tao, L. Liu, S. Oda, T. Hoshida, and J. Rasmussen, "Polarization demultiplexing based on independent component analysis in optical coherent receivers," in *Proc. Eur. Conf. Opt. Commun*, Sep. 2008, paper Mo.3.D.5.

[50] L. Liu, Z. Tao, W. Yan, S. Oda, T. Hoshida, and J. Rasmussen, "initial tap setup of constant modulus algorithm for polarization de multiplexing in optical coherent receivers," in *Proc. Opt. Fiber Commun. Conf*, Mar, 2009, paper OMT2.

[51] J. Treichler, M. Larimore, and J. Harp, "Practical blind demodulators for high-order QAM signals," *Proc. IEEE*, vol. 86, no. 10, pp. 1907–1926, Oct. 1998.

[52] I. Fatadin, D. Ives, and S. J. Savory, "Blind equalization and carrier phase recovery in a 16-QAM optical coherent system," *J. Lightw. Technol.*, vol. 27, no. 15, pp. 3042–3049, Aug. 1, 2009.

[53] P. J. Winzer and A. H. Gnauck, "112-Gb/s polarization-multiplexed 16-QAM on a 25-GHz WDM grid," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2008, paper Th.3.E.5.

[54] P. Winzer, A. Gnauck, C. Doerr, M. Magarini, and L. Buhl, "Spectrally efficient long-haul optical networking using 112-Gb/s polarization-multiplexed 16-QAM," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 547–556, Feb. 15, 2010.

[55] W. Sethares, G. Rey, and J. Johnson C.R., "Approaches to blind equalization of signals with multiple modulus," in *Proc. Int. Conf. Acoust., Speech, Signal Process., 1989 (ICASSP)*, May, vol. 2, pp. 972–975.

[56] M. Ready and R. Gooch, "Blind equalization based on radius directed adaptation," in *Proc. Int. Conf. Acoust., Speech, Signal Process., 1990 (ICASSP)*, Apr., vol. 3, pp. 1699–1702.

[57] C. Fludger, T. Duthel, D. van den Borne, C. Schulien, E.-D. Schmidt, T. Wuth, J. Geyer, E. De Man, K. Giok-Djan, and H. de Waardt, "Coherent

equalization and POLMUX-RZ-DQPSK for robust 100-GE transmission," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 64–72, Jan. 1, 2008.
[58] C. Laperle, B. Villeneuve, Z. Zhang, D. McGhan, H. Sun, and M. O'Sullivan, "WDM Performance and PMD tolerance of a coherent 40-Gbit/s dual-polarization QPSK transceiver," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 168–175, Jan. 1, 2008.
[59] O. Bertran-Pardo, J. Renaudier, G. Charlet, P. Tran, H. Mardoyan, M. Salsi, and S. Bigo, "Experimental assessment of interactions between nonlinear impairments and polarization-mode dispersion in 100-Gb/s coherent systems versus receiver complexity," *IEEE Photon. Technol. Lett.*, vol. 21, no. 1, pp. 51–53, Jan. 1, 2009.
[60] E. Ip and J. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2033–2043, Aug. 2007.
[61] B. Spinnler, "Complexity of algorithms for digital coherent receivers," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2009, paper 7.3.6.
[62] A. Vgenis, C. Petrou, C. Papadias, I. Roudas, and L. Raptis, "Nonsingular constant modulus equalizer for PDM-QPSK coherent optical receivers," *IEEE Photon. Technol. Lett.*, vol. 22, no. 1, pp. 45–47, Jan. 1, 2010.
[63] T. Oktem, A. Erdogan, and A. Demir, "Adaptive receiver structures for fiber communication systems employing polarization-division multiplexing," *J. Lightw. Technol.*, vol. 27, no. 23, pp. 5394–5404, Dec. 1, 2009.
[64] S. Haykin, "Signal processing: Where physics and mathematics meet," *IEEE Signal Process. Mag*, vol. 18, no. 4, pp. 6–7, Jul. 2001.
[65] H. Hurwitz and R. C. Jones, "A new calculus for the treatment of optical systems," *J. Opt. Soc. Amer.*, vol. 31, no. 7, pp. 493–495, 1941.
[66] M. J. D. Powell *Approximation Theory and Methods*, Cambridge, U.K. Cambridge Univ. Press, 1982.
[67] P. Prandoni and M. Vetterli, "From Lagrange to Shannon. and back: Another look at sampling [DSP Education]," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 138–144, Sep. 2009.
[68] F. M. Gardner, "Interpolation in digital modems. Part I: Fundamentals," *IEEE Trans. Commun.*, vol. 41, no. 3, pp. 501–507, Mar. 1993.
[69] L. Erup, F. M. Gardner, and R. A. Harris, "Interpolation in digital modems. Part II: Fundamentals," *IEEE Trans. Commun.*, vol. 41, no. 6, pp. 998–1008, Jun. 1993.
[70] S. H. Chang, H. S. Chung, and K. Kim, "Digital non-data-aided symbol synchronization in optical coherent intradyne reception," *Opt. Exp.*, vol. 16, no. 19, pp. 15 097–15 103, 2008.
[71] F. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers," *IEEE Trans. Commun.*, vol. COM-34, no. 5, pp. 423–429, May 1986.
[72] K. Mueller and M. Muller, "Timing recovery in digital synchronous data receivers," *IEEE Trans. Commun.*, vol. COM-24, no. 5, pp. 516–531, May 1976.
[73] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley Interscience, 2001.
[74] A. Leven, N. Kaneda, U.-V. Koc, and Y.-K. Chen, "Frequency estimation in intradyne reception," *IEEE Photon. Technol. Lett.*, vol. 19, no. 6, pp. 366–368, Mar. 15, 2007.
[75] S. Hoffmann, S. Bhandare, T. Pfau, O. Adamczyk, C. Wordehoff, R. Peveling, M. Porrmann, and R. Noe, "Frequency and phase estimation for coherent QPSK transmission with unlocked DFB lasers," *IEEE Photon. Technol. Letts*, vol. 20, no. 18, pp. 1569–1571, Sep. 15, 2008.
[76] M. Morelli and U. Mengali, "Feedforward frequency estimation for PSK: A tutorial review," *Eur. Trans. Telecommun.*, vol. 9, no. 2, pp. 103–116, 1998.
[77] M. Selmi, Y. Jaouën, and P. Ciblat, "Accurate digital frequency offset estimator for coherent PolMux QAM transmission systems," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2009, paper P3.08.
[78] K. Piyawanno, M. Kuschnerov, B. Spinnler, and B. Lankl, "Fast and accurate automatic frequency control for coherent receivers," in *Proc. Eur. Conf. Opt. Commun.*, Sep. 2009, paper 7.3.1.
[79] R. Noe, "Phase noise-tolerant synchronous QPSK/BPSK baseband-type intradyne receiver concept with feedforward carrier recovery," *J. Lightw. Technol.*, vol. 23, no. 2, pp. 802–808, Feb. 2005.
[80] E. Ip and J. Kahn, "Feedforward carrier recovery for coherent optical communications," *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2675–2692, Sep. 2007.
[81] 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. 29, no. 4, pp. 543–551, Jul. 1983.
[82] M. G. Taylor, "Phase estimation methods for optical coherent detection using digital signal processing," *J. Lightw. Technol.*, vol. 27, no. 7, pp. 901–914, Apr. 1, 2009.
[83] L. Pessoa, H. Salgado, and I. Darwazeh, "Performance evaluation of phase estimation algorithms in equalized coherent optical systems," *IEEE Photon. Technol. Lett.*, vol. 21, no. 17, pp. 1181–1183, Sep. 1, 2009.
[84] S. Hoffmann, R. Peveling, T. Pfau, O. Adamczyk, R. Eickhoff, and R. Noe, "Multiplier-free real-time phase tracking for coherent QPSK receivers," *IEEE Photon. Technol. Lett.*, vol. 21, no. 3, pp. 137–139, Feb. 1, 2009.
[85] S. Zhang, P. Kam, C. Yu, and J. Chen, "Laser linewidth tolerance of decision-aided maximum likelihood phase estimation in coherent optical M-ary PSK and QAM systems," *IEEE Photon. Technol. Lett.*, vol. 21, no. 15, pp. 1075–1077, Aug. 1, 2009.
[86] I. Fatadin, D. Ives, and S. J. Savory, "Compensation of frequency offset for differentially encoded 16- and 64-QAM in the presence of laser phase noise," *IEEE Photon. Technol. Lett.*, vol. 22, no. 3, pp. 176–178, Feb. 1, 2010.
[87] T. Pfau, S. Hoffmann, and R. Noe, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M -QAM constellations," *J. Lightw. Technol.*, vol. 27, no. 8, pp. 989–999, Apr. 15, 2009.
[88] 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. 15 718–15 727, 2008.
[89] A. P. T. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *J. Lightw. Technol.*, vol. 25, no. 10, pp. 3008–3016, 2007.
[90] I. B. Djordjevic, L. L. Minkov, L. Xu, and T. Wang, "Suppression of fiber nonlinearities and PMD in coded-modulation schemes with coherent detection by using turbo equalization," *J. Opt. Commun. Netw.*, vol. 1, no. 6, pp. 555–564, 2009.
[91] S. L. Jansen, I. Morita, T. C. W. Schenk, N. Takeda, and H. Tanaka, "Coherent optical 25.8-Gb/s OFDM transmission over 4160-km SSMF," *J. Lightw. Technol.*, vol. 26, no. 1, pp. 6–15, Jan. 2008.
[92] X. Liu, F. Buchali, and R. Tkach, "Improving the nonlinear tolerance of polarization-division-multiplexed CO-OFDM in long-haul fiber transmission," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3632–3640, Aug. 15, 2009.
[93] Y. Ma, Q. Yang, Y. Tang, S. Chen, and W. Shieh, "1-Tb/s single-channel coherent optical OFDM transmission over 600-km SSMF fiber with subwavelength bandwidth access," *Opt. Exp.*, vol. 17, no. 11, pp. 9421–9427, 2009.
[94] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," *Phys. Rev. Lett.*, vol. 101, no. 16, p. 163901, 2008.

Seb J. Savory (M'07) was born in Stirling, U.K., in 1973. He received the M.Eng., M.A., and Ph.D. degrees in engineering from the University of Cambridge, Cambridge, U.K., in 1996, 1999, and 2001, respectively, and the M.Sc. degree in mathematics from the Open University, Milton Keynes, U.K., in 2007.

His interest in optical communications began in 1991, when he joined Standard Telecommunications Laboratories, Harlow, U.K., prior to being sponsored though his undergraduate and postgraduate studies by Nortel. On completion of his Ph.D. degree in 2000, he joined Nortel's Harlow Laboratories as Senior Research Engineer, where he was engaged in research into digital signal processing and advanced optical transmission systems. In 2005, he joined the Optical Networks Group at University College London (UCL), where he held a Leverhulme Trust Early Career Fellowship from 2005–2007, being appointed as a University Lecturer in 2007. From June 2009–June 2010 he was also a Visiting Professor at the Politecnico di Torino, Italy. His research interests include digital coherent transceivers, optical transmission systems and subsystems, digital signal processing, and nonlinear systems.

Dr. Savory is a Chartered Engineer and an Associate Editor for the IEEE PHOTONICS TECHNOLOGY LETTERS. He also serves on the technical program committee for the Optical Fiber Communication Conference, the European Conference on Optical Communication, and the IEEE Photonics Society Annual Meeting.