Maximum Likelihood-Based Blind Dispersion Estimation for Coherent Optical Communication

Henk Wymeersch, *Member, IEEE,* and Pontus Johannisson

***Abstract*—Starting from the maximum likelihood criterion, we derive a novel blind chromatic dispersion (CD) estimation method in the presence of unknown data, propagation delay, polarization state, and differential group delay. By using CD estimation, electronic dispersion compensation (EDC) can be carried out without prior knowledge about the amount of accumulated CD. This adds flexibility to the EDC, which may prove valuable in reconfigurable optical networks. Using numerical simulations, we compare the suggested algorithm with a well-known CD estimation algorithm based on the constant modulus algorithm. We find that the proposed algorithm has better estimation performance and lower computational complexity. Furthermore, the impact of differential group delay is small. The derivation of the algorithm also shows the close connection between CD estimation, clock recovery, and polarization effects.**

***Index Terms*—Optical fiber communication, coherent optical communication, optical fiber dispersion.**

I. Introduction

COHERENT optical receivers must compensate for various channel impairments. For performance and flexibility reasons, many of the receiver tasks are carried out in the electrical domain through digital signal processing (DSP) [1]. The typical stages in the DSP include: (i) electronic dispersion compensation (EDC), often in the frequency domain for large amounts of accumulated chromatic dispersion (CD); (ii) clock recovery and interpolation; (iii) polarization demultiplexing and adaptive equalization; (iv) frequency offset compensation and phase synchronization; (v) data detection and decoding. While CD is a static effect, characterized by the dispersion parameter and the fiber length, the receiver must have access to this parameter in order to perform EDC. In a flexible system, where the signal path through the network may change with time, the amount of accumulated CD may not be known.

Several approaches for CD estimation have already been reported. One way is to implement *optical performance monitoring*. In [2], the residual CD and other channel parameters were estimated from the filter response in the adaptive equalizer. This method can monitor the system, but the value for the EDC is somewhat limited since CD estimation in this way relies on feedback from a later (the equalizer) to an earlier (the EDC) DSP block. CD estimation without performing the equalization was discussed in [3]. The suggested algorithm was based on the constant modulus algorithm (CMA) [4] and assumes that the sampling rate is two samples per symbol. Another CD estimation method, based on the autocorrelation of the signal spectrum was proposed in [5]–[8]. Furthermore, CD estimation based on delay-tap sampling [9], mean signal power and eigenvalue spread [10] has been investigated. All these algorithms are blind, i.e., they operate without any prior knowledge of the data. This is attractive since the algorithms are easy to integrate into the receiver and they have zero spectral efficiency penalty. The alternative to a blind algorithm is to use data-aided estimation [11], [12], at a cost of periodically inserting training sequences. While these approaches lead to practical methods, they are generally derived in an ad-hoc manner.

In this paper, we describe a novel blind CD estimation algorithm with a rigorous mathematical underpinning: starting from the maximum likelihood (ML) criterion, we are able to estimate the CD in the presence of unknown data, timing offset, differential group delay (DGD), and polarization state. The estimator operates in the frequency domain and is easy to integrate with existing frequency domain EDC. We evaluate the method by numerical simulations in terms of the estimation error variance and compare with a well-known method from [3]. Our results indicate that the ML-based estimator achieves lower error variance at a smaller computational cost. We also find that the performance penalty from DGD is very limited for any amount of DGD and can, if necessary, be reduced further at a marginal increase in computational complexity.

Although we include many of the most important signal transmission impairments, there are even more physical phenomena that can be studied in this context. Examples of this are the Kerr nonlinearity and even more general polarization effects, such as polarization-dependent loss and polarization-mode dispersion. However, such an investigation is outside the scope of this work and we leave this for a later study.

Notation: $\mathbb{E}_x\{\cdot\}$ denotes the expectation operator with respect to the random variable x; the probability density function of $\mathbf{x}$ is written $p(\mathbf{x})$; vectors are denoted by bold letters; quantities in the frequency domain are generally denoted by capital letters; $\|\mathbf{x}\|$ denotes vector norm, i.e., $\|\mathbf{x}\| = \sqrt{\mathbf{x}^{\mathrm{H}}\mathbf{x}}$; the identity matrix is denoted by $\mathbf{I}$; $\Re(x)$ extracts the real part of the complex number x; for any real-valued function $f(\cdot)$, $\arg\max_x \max_y f(x, y)$ returns the value of x corresponding to the global maximum of $f(\cdot)$; the component in the x polarization of a vector-signal $\mathbf{y}(t)$ will be written as $y^x(t)$.

Henk Wymeersch is with the Communication Systems Group, Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. Pontus Johannisson is with the Photonics Laboratory, Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. Email: {henkw, pontus.johannisson}@chalmers.se.

II. Problem Statement

We consider a coherent optical communication system using polarization multiplexing. The transmitted signal can be written as

$$\mathbf{s}(t) = \sum_{k=-\infty}^{\infty} \mathbf{a}_k p(t - kT), \tag{1}$$

where $\mathbf{a}_k = [a_k^x, a_k^y]^{\mathrm{T}}$ is the complex 2×1 symbol vector for symbol slot k, $p(t)$ is a unit-energy transmit pulse, and $1/T$ is the symbol rate. We assume that the data symbols have the following properties: $\mathbb{E}\{\mathbf{a}_k\} = \mathbf{0}$, $\mathbb{E}\{\mathbf{a}_k \mathbf{a}_l^{\mathrm{T}}\} = \mathbf{0}$, $\forall k, l$, and $\mathbb{E}\{\mathbf{a}_k \mathbf{a}_l^{\mathrm{H}}\} = E_s \delta_{kl} \mathbf{I}$, where E_s is the energy per symbol. These assumptions hold for all M-ary quadrature amplitude modulation formats and M-ary phase-shift keying modulation formats for $M > 2$. Notable exceptions include on–off keying and binary phase-shift keying.

The phase and polarization state of the received signal are unknown, as is the differential group delay (DGD) along an unknown axis. We model these effects through the Jones matrix

$$\mathbf{T}(f) = \mathbf{V}\mathbf{D}(f)\mathbf{U} = \mathbf{V} \begin{pmatrix} e^{-j\pi f T_{\mathrm{DGD}}} & 0 \\ 0 & e^{j\pi f T_{\mathrm{DGD}}} \end{pmatrix} \mathbf{U}, \tag{2}$$

where $\mathbf{U}$ and $\mathbf{V}$ are general unknown unitary Jones matrices that can change the polarization to any possible state and T_{DGD} is the total difference in propagation times along the slow and fast axes, respectively. The received electrical signal can be expressed in the frequency domain as

$$\begin{aligned} \mathbf{R}(f) &= H(f) e^{-j2\pi f \tau} \mathbf{T}(f) \mathbf{S}(f) + \mathbf{N}(f) \\ &\equiv \mathbf{X}(f) + \mathbf{N}(f), \end{aligned} \tag{3}$$

where $H(f)$ is the transfer function of the CD, τ is an unknown propagation delay, $\mathbf{S}(f)$ is the Fourier transform of the transmitted signal, and $\mathbf{N}(f)$ is complex additive white Gaussian noise (AWGN) with power spectral density $N_0/2$ per each of the four real dimensions. The CD all-pass filter is described by

$$H(f) = \exp\left(-j \frac{\pi f^2 \lambda^2}{c} \int_0^L D(z) \mathrm{d}z \right), \tag{4}$$

where λ is the carrier wavelength, c is the speed of light, $D(z)$ is the dispersion parameter, and L is the total system length. Using the expression for the group delay [13, p. 39]

$$\Delta T = D L \Delta\lambda, \tag{5}$$

we introduce

$$\eta \equiv \frac{\Delta T}{T} = \frac{\Delta \lambda}{T} \int_0^L D(z) \mathrm{d}z = -\frac{\lambda^2}{c T^2} \int_0^L D(z) \mathrm{d}z, \tag{6}$$

where we used the symbol rate to find a representative bandwidth $\Delta\lambda$ for the channel. The transfer function (4) can then be written $H(f) = \exp(j\pi\eta T^2 f^2)$. In the case $|\eta| \gg 1$, we can directly interpret $|\eta|$ as the signal pulse broadening measured in symbol slots. Our goal is to set up a structure of the form shown in Fig. 1, where we estimate η based on $\mathbf{r}(t)$, the observed time-domain counterpart of $\mathbf{R}(f)$, without knowledge of τ, $\mathbf{T}$, or $\mathbf{a}_k$.

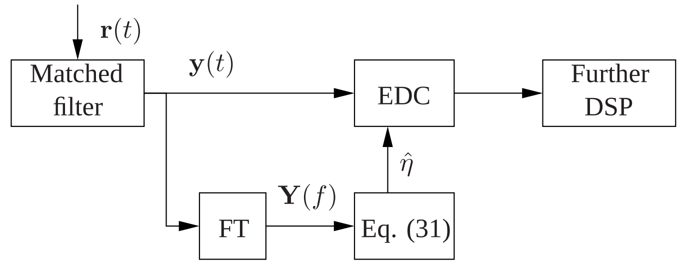

Figure 1. Block diagram of the proposed CD estimator. The incoming signal is filtered, Fourier transformed (FT), and given to the estimator, which is stated later as (31). The estimated value is used in the EDC and the "further DSP" includes, e.g., adaptive equalization, phase synchronization, and data detection.

III. Proposed CD Estimator

A. The ML Criterion

ML estimators are widely used in estimation problems, as they have a number of desirable asymptotic properties, such as consistency and efficiency. According to the ML criterion, the optimal estimate of an unknown parameter $\mathbf{x}$ in the presence of a random nuisance parameter $\mathbf{z}$ from an observation $\mathbb{r}$ is

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\max_{\mathbf{x}} p(\mathbb{r}|\mathbf{x}) = \arg\max_{\mathbf{x}} \int p(\mathbb{r}, \mathbf{z}|\mathbf{x}) \mathrm{d}\mathbf{z} \\ &= \arg\max_{\mathbf{x}} \int p(\mathbb{r}|\mathbf{x}, \mathbf{z}) p(\mathbf{z}|\mathbf{x}) \mathrm{d}\mathbf{z}. \end{aligned} \tag{7}$$

In our case, we make the following associations: $\mathbf{x} \leftrightarrow [\eta, \tau, \mathbf{T}]$, $\mathbf{z} \leftrightarrow \mathbf{a}$. For a more general description of ML estimation, see [14, Section 12.2].

B. CD Estimation

Using the fact that the noise is AWGN and suppressing the infinite limits to make the notation compact, we obtain

$$\begin{aligned} p(\mathbb{r}|\eta, \tau, \mathbf{T}, \mathbf{a}) &\propto \exp\left[-\frac{1}{N_0} \int \|\mathbf{r}(t) - \mathbf{x}(t)\|^2 \mathrm{d}t \right] \\ &= \exp\left[-\frac{1}{N_0} \int \|\mathbf{R}(f) - \mathbf{X}(f)\|^2 \mathrm{d}f \right], \end{aligned} \tag{8}$$

where we used Parseval's theorem to write the likelihood in the Fourier domain, and denoted a vector representation of $\mathbf{r}(t)$ by $\mathbb{r}$. Expanding the vector norm we obtain

$$p(\mathbb{r}|\eta, \tau, \mathbf{T}, \mathbf{a}) \propto \exp\left[\frac{2}{N_0} \Re \int \mathbf{R}^{\mathrm{H}}(f) \mathbf{X}(f) \mathrm{d}f \right], \tag{9}$$

where we have removed the additive terms that do not depend on η. We Taylor expand to second order according to $\exp(x) \approx 1 + x + x^2/2$ to obtain

$$\begin{aligned} p(\mathbb{r}|\eta, \tau, \mathbf{T}, \mathbf{a}) \propto 1 &+ \frac{2}{N_0} \Re \int \mathbf{R}^{\mathrm{H}}(f) \mathbf{X}(f) \mathrm{d}f \\ &+ \frac{2}{N_0^2} \left[\Re \int \mathbf{R}^{\mathrm{H}}(f) \mathbf{X}(f) \mathrm{d}f \right]^2. \end{aligned} \tag{10}$$

It should be noticed that the Taylor expansion will yield suboptimal results at a high signal-to-noise ratio (SNR). The first (constant) term in (10) can be dropped, as it will not affect the ML estimate.

1) Expectation over the Data: We now take the expectation of (10) with respect to the data symbols $\mathbf{a}_k$. For the second term in (10), we have

$$\mathbb{E}_{\mathbf{a}}\left\{\frac{2}{N_0}\Re\int \mathbf{R}^{\mathrm{H}}\mathbf{X}\mathrm{d}f\right\} = \frac{2}{N_0}\Re\int \mathbf{R}^{\mathrm{H}}\mathbb{E}_{\mathbf{a}}\left\{\mathbf{X}\right\}\mathrm{d}f, \quad (11)$$

in which

$$\mathbb{E}_{\mathbf{a}}\left\{\mathbf{X}\right\} = \sum_k e^{j\pi\eta T^2 f^2}\mathbf{T}\mathbb{E}_{\mathbf{a}_k}\{\mathbf{a}_k\}e^{-j2\pi(kT+\tau)f}P(f). \quad (12)$$

As discussed above, $\mathbb{E}_{\mathbf{a}_k}\{\mathbf{a}_k\} = \mathbf{0}$ and the second term in (10) is therefore zero. Using the relation $2\Re(x) = x + x^*$ and neglecting the irrelevant positive multiplicative factor, the third term in (10) can be written as

$$\left[\int \mathbf{R}^{\mathrm{H}}(f)\mathbf{X}(f)\mathrm{d}f\right]^2 + \left[\int \mathbf{X}^{\mathrm{H}}(f)\mathbf{R}(f)\mathrm{d}f\right]^2 + 2\int \mathbf{R}^{\mathrm{H}}(f)\mathbf{X}(f)\mathrm{d}f\int \mathbf{X}^{\mathrm{H}}(v)\mathbf{R}(v)\mathrm{d}v. \quad (13)$$

Using (3) to substitute $\mathbf{X}(f)$ and taking the expectation with respect to $\mathbf{a}_k$, the first term of (13) will contain terms of the form $\mathbb{E}_{\mathbf{a}}\{\mathbf{a}_k\mathbf{a}_l^{\mathrm{T}}\}$. Similarly, the second term will contain terms of the form $\mathbb{E}_{\mathbf{a}}\{\mathbf{a}_k^*\mathbf{a}_l^{\mathrm{H}}\}$. These two terms are both zero due to the assumptions about the constellation made above. Finally, the third term in (13) will contain terms of the form $\mathbb{E}_{\mathbf{a}}\{\mathbf{a}_k\mathbf{a}_l^{\mathrm{H}}\} = E_s\delta_{kl}\mathbf{I}$. In summary, we find that

$$p(\mathbb{r}|\eta,\tau,\mathbf{T}) = \sum_{\mathbf{a}} p(\mathbb{r}|\eta,\tau,\mathbf{T},\mathbf{a})p(\mathbf{a})$$
$$\propto \sum_k \iint \mathbf{R}^{\mathrm{H}}(f)\mathbf{T}(f)\mathbf{T}^{\mathrm{H}}(v)\mathbf{R}(v)P(f)P^*(v)$$
$$\times\, e^{-j2\pi(kT+\tau)f}e^{j\pi\eta T^2 f^2}e^{j2\pi(kT+\tau)v}e^{-j\pi\eta T^2 v^2}\mathrm{d}f\mathrm{d}v, \quad (14)$$

where we have dropped additive constants and positive multiplicative factors. Observe that (14) does not depend on the data.

2) Interpretation of the Likelihood Function: We introduce

$$\mathbf{Z}(f;\eta,\mathbf{T}) = e^{-j\pi\eta T^2 f^2}P^*(f)\mathbf{T}^{\mathrm{H}}(f)\mathbf{R}(f), \quad (15)$$

with the corresponding time domain signal $\mathbf{z}(t;\eta,\mathbf{T})$, which we interpret as the received signal after a matched filter (with frequency response $P^*(f)$), polarization demultiplexing and DGD correction, and EDC with a candidate value for η. We can then write

$$p(\mathbb{r}|\eta,\tau,\mathbf{T}) \propto \sum_k \left\|\int \mathbf{Z}(f;\eta,\mathbf{T})e^{j2\pi(kT+\tau)f}\mathrm{d}f\right\|^2, \quad (16)$$

or

$$p(\mathbb{r}|\eta,\tau,\mathbf{T}) \propto \sum_k \|\mathbf{z}(kT+\tau;\eta,\mathbf{T})\|^2. \quad (17)$$

Hence, the likelihood function (14) has a direct and very intuitive interpretation: *η, τ, and $\mathbf{T}$ should be chosen such that, when using these values to do EDC, compensation for the polarization and DGD effects, and sampling at times $kT+\tau$, then the sum of the symbol power over all symbols and both polarizations is maximized.* In words, this can be loosely described as “maximizing the eye opening”. The ML estimates of η, τ, and $\mathbf{T}$ are found by solving the optimization problem

$$[\hat{\eta},\hat{\tau},\hat{\mathbf{T}}] = \arg\max_{\eta,\tau,\mathbf{T}} p(\mathbf{r}|\eta,\tau,\mathbf{T}). \quad (18)$$

3) Removal of the Dependency on τ: It is clear that a direct optimization of (18) calls for joint estimation of η, τ, and $\mathbf{T}$, which may be impossible due to the high computational complexity. To avoid this, we rewrite (16) as

$$p(\mathbb{r}|\eta,\tau,\mathbf{T}) \propto \quad (19)$$
$$\iint \mathbf{Z}^{\mathrm{H}}(f;\eta,\mathbf{T})\mathbf{Z}(v;\eta,\mathbf{T})e^{j2\pi\tau(v-f)}\sum_k e^{j2\pi kT(v-f)}\mathrm{d}v\mathrm{d}f,$$

and use

$$\sum_{k=-\infty}^{\infty} e^{j2\pi kT(v-f)} = \frac{1}{T}\sum_{k=-\infty}^{\infty}\delta(v-(f+k/T)) \quad (20)$$

to perform the integration over v. We then obtain

$$p(\mathbb{r}|\eta,\tau,\mathbf{T}) \propto$$
$$\sum_k e^{j2\pi\tau k/T}\int \mathbf{Z}^{\mathrm{H}}(f;\eta,\mathbf{T})\mathbf{Z}(f+k/T;\eta,\mathbf{T})\mathrm{d}f. \quad (21)$$

Examining this sum, we find that the term corresponding to $k=0$ is the total signal energy, which is not affected by the choice of η, τ, or $\mathbf{T}$. Consequently, we can drop this term. Furthermore, we find that the complex conjugate of term k is equal to the term $-k$, allowing us to reformulate the summation as

$$p(\mathbb{r}|\eta,\tau,\mathbf{T}) \propto$$
$$\Re\sum_{k=1}^{\infty} e^{j2\pi\tau k/T}\int \mathbf{Z}^{\mathrm{H}}(f;\eta,\mathbf{T})\mathbf{Z}(f+k/T;\eta,\mathbf{T})\mathrm{d}f. \quad (22)$$

In this infinite sum, terms corresponding to $k>1$ will be small, as the spectra $\mathbf{Z}^{\mathrm{H}}(f;\eta,\mathbf{T})$ and $\mathbf{Z}(f+k/T;\eta,\mathbf{T})$ will only have a small overlap. Hence, retaining only the term corresponding to $k=1$, we obtain

$$[\hat{\eta},\hat{\tau},\hat{\mathbf{T}}] = \quad (23)$$
$$\arg\max_{\eta,\tau,\mathbf{T}} \Re\left[e^{j2\pi\tau/T}\int \mathbf{Z}^{\mathrm{H}}(f;\eta,\mathbf{T})\mathbf{Z}(f+1/T;\eta,\mathbf{T})\mathrm{d}f\right].$$

This formulation allows us to decompose the optimization to

$$[\hat{\eta},\hat{\mathbf{T}}] = \arg\max_{\eta,\mathbf{T}}\left|\int \mathbf{Z}^{\mathrm{H}}(f;\eta,\mathbf{T})\mathbf{Z}(f+1/T;\eta,\mathbf{T})\mathrm{d}f\right| \quad (24)$$

and

$$\hat{\tau} = -\frac{T}{2\pi}\angle\left[\int \mathbf{Z}^{\mathrm{H}}(f;\hat{\eta},\hat{\mathbf{T}})\mathbf{Z}(f+1/T;\hat{\eta},\hat{\mathbf{T}})\mathrm{d}f\right], \quad (25)$$

where $\angle x$ denotes the phase of the complex variable x. The estimate $\hat{\tau}$ in (25) corresponds to the well-known Oerder and Meyr estimator [15]. This highlights the close connection between CD estimation, clock recovery, and polarization effects. Note that when η, τ, and $\mathbf{T}$ have been estimated, then the EDC and the interpolation can be carried out in the most convenient order. This offers a potential re-ordering of receiver components.

4) Final CD Estimator: We choose to formulate the final result (24) in terms of the received signal after matched filtering only. Thus, we introduce $\mathbf{Y}(f) = P^*(f)\mathbf{R}(f)$, see Fig. 1. The estimator becomes

$$\hat{\eta} = \arg\max_{\eta}\max_{\mathbf{M}}\left|\int \mathbf{Y}^{\mathrm{H}}(f)\mathbf{M}\mathbf{Y}(f+1/T)e^{-j2\pi\eta Tf}\mathrm{d}f\right|, \tag{26}$$

where

$$\mathbf{M} = \mathbf{V}\mathbf{D}(f)\mathbf{U}\mathbf{U}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}(f+1/T)\mathbf{V}^{\mathrm{H}}. \tag{27}$$

Since $\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{I}$, we find

$$\mathbf{M} = \mathbf{V}\begin{pmatrix} e^{j\pi T_{\mathrm{DGD}}/T} & 0 \\ 0 & e^{-j\pi T_{\mathrm{DGD}}/T}\end{pmatrix}\mathbf{V}^{\mathrm{H}}, \tag{28}$$

which does not depend on f and shows explicitly that (26) is periodic in T_{DGD} with period T. It is now possible to estimate both η and $\mathbf{M}$, but the aim here is only to estimate η. To minimize the computational complexity of the CD estimation, we should therefore test only a minimal set of matrices, $\mathcal{S}$, selected in such a way as to achieve sufficiently high accuracy in the estimation of η. In the absence of DGD, we have $\mathbf{M} = \mathbf{I}$. If $T_{\mathrm{DGD}} \neq 0$, then $\mathbf{M}$ depends on T_{DGD} and $\mathbf{V}$ but we notice that $\mathbf{M}$ is unitary. Using [16, Eq. (4.4)] it can therefore be written on the form

$$\mathbf{I}\cos(\varphi/2) - j\hat{\mathbf{r}}\cdot\vec{\boldsymbol{\sigma}}\sin(\varphi/2), \tag{29}$$

where φ and $\hat{\mathbf{r}}$ are the rotation angle and normalized axis, respectively, and $\vec{\boldsymbol{\sigma}}$ is the Pauli spin vector. This shows that every matrix $\mathbf{M}$ can be written as a linear combination of the identity and the Pauli spin matrices, i.e., the set $\{\mathbf{I}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3\}$. We suggest to choose $\mathcal{S}$ as this set of matrices. However, since $\mathbf{M}$ can be multiplied by j without affecting the estimate (26), we can select an equally well performing set as

$$\mathcal{S} = \left\{\begin{pmatrix} 1 & 0 \\ 0 & \pm 1\end{pmatrix}, \begin{pmatrix} 0 & 1 \\ \pm 1 & 0\end{pmatrix}\right\}. \tag{30}$$

To motivate this set intuitively, we first notice that since $\mathbf{D}$ is periodic in T_{DGD} with period T, we expect that the case $T_{\mathrm{DGD}} = T/2$ would be a worst case. By using (29) to construct $\mathbf{V}$ matrices corresponding to $m\pi/2$ rotations, where m is an arbitrary integer, around any of the three axes in Stokes space, we find that all these cases are handled exactly by the set $\mathcal{S}$ for $T_{\mathrm{DGD}} = T/2$. This shows that the set $\mathcal{S}$ includes more cases than what is immediately obvious. The final estimator, to be used where indicated in Fig. 1, becomes

$$\hat{\eta} = \arg\max_{\eta}\max_{\mathbf{M}\in\mathcal{S}}\left|\int \mathbf{Y}^{\mathrm{H}}(f)\mathbf{M}\mathbf{Y}(f+1/T)e^{-j2\pi\eta Tf}\mathrm{d}f\right|. \tag{31}$$

As described in Section IV-C, this selection of the set $\mathcal{S}$ leads to a small performance penalty for any amount of DGD. However, the penalty can, if necessary, be made arbitrarily small by increasing the number of matrices in the set $\mathcal{S}$. The increase of the computational complexity is marginal as a larger number of $\mathbf{M}$ matrices is included, which is seen by writing (31) in terms of $\int (Y^a(f))^* Y^b(f+1/T)e^{-j2\pi\eta Tf}\mathrm{d}f$ for $a, b \in \{x, y\}$ and the matrix elements of $\mathbf{M}$. Then, for a given value of η, the four integrals can be calculated first, and the results are used to form a linear combination with the elements of the matrix $\mathbf{M}$.

We notice that the final result (31) bears resemblance to the estimator suggested in [6], which was not derived directly from the ML criterion. However, some differences exist, e.g., (i) [6, Eq. (1)] is for a scalar field and no explicit algorithm is given for a polarization-multiplexed signal; (ii) while DGD and PMD are mentioned in [6], no detailed method for handling DGD is given; (iii) The spectral shift is $\pm 1/T$ in [6], while it is $1/T$ in (31); (iv) two different cost functions are suggested in [6], while the ML approach results in one single algorithm.

C. Practical Considerations

In a DSP implementation, we would sample the uncompensated matched filter output $\mathbf{y}(t)$ asynchronously at a rate $1/T_s$ during an observation window of N symbols (or $N_{\mathrm{obs}} = NT/T_s$ samples), and apply a fast Fourier transform to obtain a vector $\mathbf{Y}$ with frequency resolution $\delta f = 1/(NT)$. Then we find that

$$\hat{\eta} = \arg\max_{\eta}\max_{\mathbf{M}\in\mathcal{S}}\left|\sum_{k=0}^{N_{\mathrm{obs}}-1}\mathbf{Y}_k^{\mathrm{H}}\mathbf{M}\mathbf{Y}_{k+N}e^{-j2\pi\eta Tk\delta f}\right|. \tag{32}$$

As a side-effect, we notice that the objective function (32) is now periodic with period N. This implies that the *a priori* uncertainty in η must not exceed N. In the discrete domain, (32) also places a restriction on the sampling rate at the receiver: assuming the bandwidth of $\mathbf{Y}(f)$ is approximately $[-B, +B]$, then to avoid aliasing, the sampling rate $1/T_s$ should satisfy $1/T_s > 2B + 1/T$. Finally, we mention that the complexity of the estimator (32) scales as $N_{\mathrm{obs}}(\log(N_{\mathrm{obs}}) + N_{\mathrm{eval}})$, where N_{eval} is the number of values of η for which the objective function is evaluated.

IV. Numerical Simulations

A. Simulation Setup

In order to evaluate the suggested algorithm numerically, we use Monte Carlo simulations, and average our results over 1000 runs. We consider a 28 Gbaud system using polarization-multiplexed quadrature phase-shift keying transmission using RZ pulses with 50 % duty cycle. The polarization mixing matrices $\mathbf{U}$ and $\mathbf{V}$, and delay τ are drawn uniformly from the set of 2×2 unitary matrices, and $[0, 10T]$, respectively. To allow a fair evaluation with a benchmark algorithm (see below), the received signal is sampled without knowledge of τ at two samples per symbol, and then upsampled in DSP to four samples per symbol to fulfill the sampling rate requirement in Section III-C. We assume $D(z) = 17$ ps/(nm km), $\lambda = 1550$ nm, and $L \in \{300\text{ km}, 3000\text{ km}\}$, corresponding to an accumulated dispersion of 5100 ps/nm and 51000 ps/nm, respectively (equivalently, $\eta \in \{-32.04, -320.43\}$). Note that for $N = 512$, the ambiguity period of the ML-based estimator is 81491 ps/nm. The optimization problem (31) is first solved by a coarse search over the interval $[0, 10000]$ ps/nm for $L = 300$ km, and over $[50000, 60000]$ ps/nm for $L = 3000$ km, each time with a resolution of 200 ps/nm. The shape of the cost

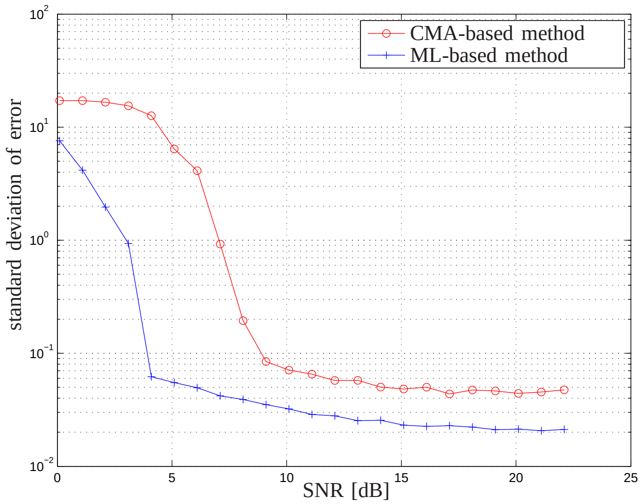

Figure 2. Standard deviation of the estimation error of η for the 300 km link and an observation duration of 512 symbols.

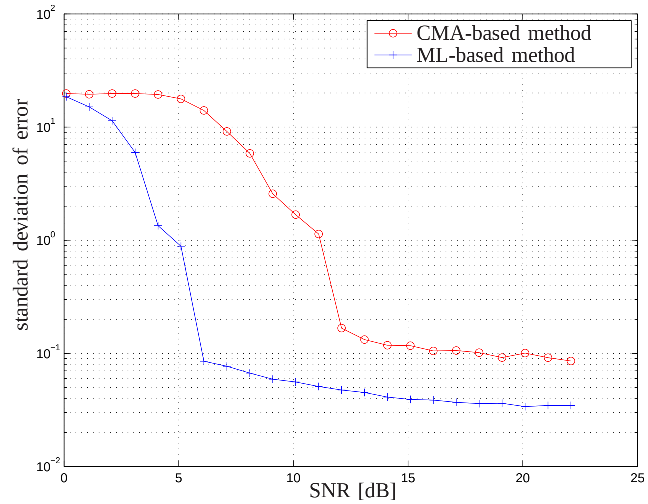

Figure 3. Standard deviation of the estimation error of η for the 3000 km link and an observation duration of 512 symbols.

function is similar to what has been found for other estimation methods; a narrow peak at the true value and smaller peaks elsewhere. (See, e.g., [7, Fig. 4].) A fine search is then performed using the `fminsearch` function in MATLAB. The SNR at the receiver is varied from 0 dB to 21 dB. The DGD parameter T_{DGD} is varied between 0 ps and $T = 35.72$ ps.

B. Benchmark Algorithm

We compare the proposed ML-based method with a competing estimator also operating at two samples per symbol. As described in the introduction, many algorithms have been suggested and it is not easy to select a benchmark algorithm. We choose to use the CMA-based algorithm from [3]. (Notice that this is not the CMA commonly used for polarization demultiplexing.) The reasons for this selection include that this algorithm has good performance (see [7]) and is described in detail for a polarization-multiplexed signal. We first normalize the energy per symbol and polarization to one. Then, for every candidate value of η, we apply matched filtering and EDC. This leads to samples $z^x_{k,\text{odd}}(\eta)$ and $z^y_{k,\text{odd}}(\eta)$ for the odd samples, and $z^x_{k,\text{even}}(\eta)$ and $z^y_{k,\text{even}}(\eta)$ for the even samples, $k = 1, \dots, N$. Let $P^x_{\text{odd}}(\eta)$ be the average energy of the odd samples in the x polarization, and similarly we introduce $P^y_{\text{odd}}(\eta)$, $P^x_{\text{even}}(\eta)$, and $P^y_{\text{even}}(\eta)$. Denoting $Q^x(\eta) = P^x_{\text{even}}(\eta)/P^x_{\text{odd}}(\eta)$ and $Q^y(\eta) = P^y_{\text{even}}(\eta)/P^y_{\text{odd}}(\eta)$, we compute $R^x_{\text{odd}}(\eta)$ and $R^x_{\text{even}}(\eta)$ (and similarly $R^y_{\text{odd}}(\eta)$ and $R^y_{\text{even}}(\eta)$) as

$$[R^x_{\text{odd}}(\eta)\, R^x_{\text{even}}(\eta)] = \begin{cases} [R_a\, R_c] & Q^x(\eta) > \xi \\ [R_b\, R_b] & \xi^{-1} < Q^x(\eta) < \xi \\ [R_c\, R_a] & Q^x(\eta) < \xi^{-1}, \end{cases} \quad (33)$$

where R_a, R_b, R_c, and ξ are optimized empirically. We use the values from [3], i.e., $R_a = 0.6$, $R_b = 1.5$, $R_c = 2$, and $\xi = 1.25$. Finally the cost function is

$$J(\eta) = \sum_{k=0}^{N-1} \sum_{\substack{s\in\\ \{\text{odd,even}\}}} \sum_{\substack{p\in\\ \{x,y\}}} \left| |z^p_{k,s}(\eta)|^2 - R^p_s(\eta) \right|, \quad (34)$$

which is to be minimized with respect to η. This method has a computational complexity that scales as $N_{\text{eval}} N_{\text{obs}} \log(N_{\text{obs}})$, which is substantially higher than what was found above for the proposed method. The reason for this is that EDC is carried out for each candidate value and the cost function is then evaluated in the time domain.

C. Results and Discussion

We first evaluate the performance of the two estimators through Monte Carlo simulations over 1000 runs, in a case of negligible DGD, i.e., $T_{\text{DGD}} = 0$ and $\mathbf{M} = \mathbf{I}$. The performance norm used is the standard deviation of the estimation error $\hat{\eta} - \eta$. We recall from the definition of η that the estimation error is directly related to the residual symbol broadening after EDC, expressed in symbol slots. Hence, we aim for a negligible residual broadening by demanding that the error standard deviation is below 1. The results for the 300 km and 3000 km link are shown in Figs. 2 and 3, respectively, assuming $N = 512$. We observe that the ML-based method significantly outperforms the CMA-based method in both cases, especially at low SNR values. For the 3000 km link, the ML-based method obtains a standard deviation below 1 for an SNR below 5 dB, while the CMA-based method required an SNR above 10 dB to obtain the same performance. For low SNR, the performance is mainly dominated by outliers. Both estimators exhibit a flooring phenomenon where the performance improves only slowly with increasing SNR. In this parameter regime, the performance is limited by the observation length instead of the SNR.

A different view is provided by Fig. 4, showing the performance at a fixed SNR of 9.8 dB, corresponding to an uncoded bit error rate of 10^{-3}, as a function of the observation duration N. For the 3000 km link, the ML-based estimator requires 512 symbols to obtain good estimates, while the CMA-based estimator requires around 1024 symbols to attain similar performance. For the 300 km link, both estimators can cope with shorter observation durations, but the ML-based estimator exhibits significantly better performance.

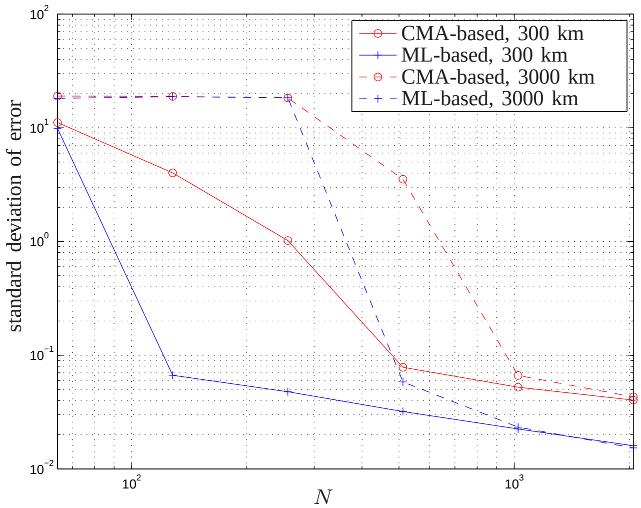

Figure 4. Standard deviation of the estimation error of η for different observation durations, N, expressed in symbols. The SNR is 9.8 dB.

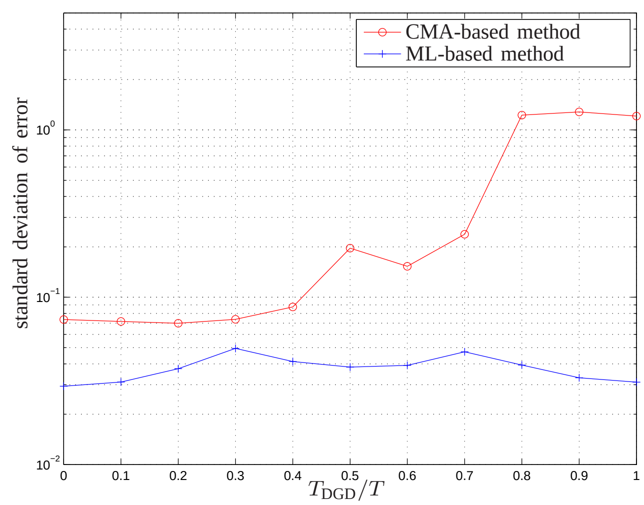

Figure 5. Standard deviation of the estimation error of η for different DGD values, T_{DGD}/T, for a 300 km link. The SNR is 9.8 dB and $N = 512$.

Finally, we investigate the impact of DGD on both estimators. We again consider a fixed SNR of 9.8 dB, and set $L = 300$ km and $N = 512$. The standard deviation of the estimation error is shown in Fig. 5. We observe that the ML-based estimator suffers only from a minor degradation with increased DGD, and that the performance is symmetric around $T_{\text{DGD}} = T/2$. The largest degradations occur for $T_{\text{DGD}} \approx 0.3T$ and $T_{\text{DGD}} \approx 0.7T$, due to the fact that the set $\mathcal{S}$ best compensates for the effect of the DGD in the cases $T_{\text{DGD}} \approx 0$ and $T_{\text{DGD}} \approx T/2$. As discussed above, the accuracy of the CD estimation can, if necessary, be further improved by expanding the set $\mathcal{S}$, defined in (30). The CMA-based method incurs a degradation with increasing DGD and this effect turns out to be mainly due to outliers in the CD estimates.

V. Conclusions

We have presented a novel blind CD estimation method derived from the ML criterion. The algorithm is able to accurately estimate the accumulated CD without knowledge of the data, propagation delay, polarization state, or differential group delay. Using numerical simulations, we have compared with a CMA-based alternative and found that the ML-based algorithm exhibits better estimation accuracy and reduced sensitivity to DGD at a lower computational complexity. The derivation of the estimator has also shown the close connection between dispersion estimation and clock recovery.

Acknowledgment

We gratefully acknowledge the inspiring discussions with Robert Borkowski and Darko Zibar, both at DTU Fotonik, Technical University of Denmark. We also want to thank Magnus Karlsson, Photonics Laboratory, Chalmers University of Technology, for helpful discussions.

References

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

[2] F. N. 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. 2009.

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

[4] D. N. 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.

[5] F. N. Hauske, C. Xie, Z. Zhang, C. Li, L. Li, and Q. Xiong, "Frequency domain chromatic dispersion estimation," in *Optical Fiber Communication Conference (OFC)*, 2010, p. JThA11.

[6] F. N. Hauske, Z. Zhang, C. Li, C. Xie, and Q. Xiong, "Precise, robust and least complexity CD estimation," in *Optical Fiber Communication Conference (OFC)*, 2011, p. JWA32.

[7] R. A. Soriano, F. N. Hauske, N. G. Gonzalez, Z. Zhang, Y. Ye, and I. T. Monroy, "Chromatic dispersion estimation in digital coherent receivers," *J. Lightw. Technol.*, vol. 29, no. 11, pp. 1627–1637, June 2011.

[8] Q. Sui, A. P. T. Lau, and C. Lu, "Fast and robust chromatic dispersion estimation using auto-correlation of signal power waveform for DSP based-coherent systems," in *Optical Fiber Communication Conference (OFC)*, 2009, p. OW4G.3.

[9] D. Wang, C. Lu, A. P. T. Lau, Member, and S. He, "Adaptive chromatic dispersion compensation for coherent communication systems using delay-tap sampling technique," *IEEE Photon. Technol. Lett.*, vol. 23, no. 14, pp. 1016–1018, July 2011.

[10] R. Borkowski, X. Zhang, D. Zibar, R. Younce, and I. T. Monroy, "Experimental demonstration of adaptive digital monitoring and compensation of chromatic dispersion for coherent DP-QPSK receiver," *Opt. Express*, vol. 19, no. 26, pp. B728–B735, Dec. 2011.

[11] F. Pittalà, F. N. Hauske, Y. Ye, N. G. Gonzalez, and I. T. Monroy, "Fast and robust CD and DGD estimation based on data-aided channel estimation," in *International Conference on Transparent Optical Networks (ICTON)*, 2011, p. We.D1.5.

[12] C. Do, A. V. Tran, and D. F. Hewitt, "Chromatic dispersion estimation based on complementary Golay sequences for 80 Gb/s QPSK single-carrier system with frequency domain equalization," in *Australasian Telecommunication Networks and Applications Conference (ATNAC)*, 2011, p. VIII:4.

[13] G. P. Agrawal, *Fiber-Optic Communication Systems*, 4th ed. John Wiley & Sons, 2010.

[14] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*. John Wiley & Sons, 1998.

[15] M. Oerder and H. Meyr, "Digital filter and square timing recovery," *IEEE Trans. Commun.*, vol. 36, no. 5, pp. 605–612, May 1988.

[16] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4541–4550, Apr. 2000.

Henk Wymeersch is Assistant Professor with the Department of Signals and Systems at Chalmers University of Technology, Sweden. Prior to joining Chalmers, he was a Postdoctoral Associate with the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT). Henk Wymeersch obtained the Ph.D. degree in electrical engineering/applied sciences in 2005 from Ghent University, Belgium. In 2006, he won the Alcatel Bell Scientific Award for his Ph.D. thesis. He is a member of the IEEE, Associate Editor for IEEE Communication Letters, for the Journal of Computer Systems, Networks, and Communications, and author of Iterative Receiver Design (Cambridge University Press, August 2007). His research interests include algorithm design for wireless transmission, coherent optical communications, and indoor navigation.

Pontus Johannisson received his Ph.D. degree from Chalmers University of Technology, Göteborg, Sweden, in 2006. His thesis was focused on nonlinear intrachannel signal impairments in optical fiber communications systems. In 2006, he joined the research institute IMEGO in Göteborg, Sweden, where he worked with digital signal processing for inertial navigation with MEMS-based accelerometers and gyroscopes. In 2009, he joined the Photonics Laboratory, Chalmers University of Technology, where he currently holds a position as Assistant Professor. His research interests include nonlinear effects in optical fibers and digital signal processing in coherent optical receivers.