Block-wise Digital Signal Processing for PolMux QAM/PSK Optical Coherent Systems

M. Selmi, C. Gosset, M. Noelle, P. Ciblat, Y. Jaouën

***Abstract*—In polarization multiplexing based coherent optical transmission, two main kinds of impairements have to be counteracted: *i)* the inter-symbol interference generated by the chromatic dispersion and the polarization mode dispersion, and *ii)* the frequency offset. Usually adaptive approaches are carried out to mitigate them. Since the channel is very slowly time-varying, we propose to combat these impairements by using block-wise methods. Therefore we introduce two new algorithms: the first one is a block-wise version of blind time equalizer (such as CMA), and the second one estimates the frequency offset in block-wise way. These algorithms are suitable for PSK and QAM constellations. By simulation investigations, we show that they outperform the standard approach in terms of convergence speed only at a moderate expense of computational load. We also experimentally evaluate their performance using 8-PSK real data traces and off-line processing which takes into account other physical impairments such as phase noise and non-linear effects.**

***Index Terms*—optical coherent communications, block processing, blind equalization, frequency offset estimation, constant phase estimation, constant modulus algorithm, decision-directed algorithm, optimal step-size gradient algorithm, bursty transmission.**

I. Introduction

COHERENT detection combined with multilevel modulation such as M-ary quadrature modulation (M-QAM) formats are one of the most relevant techniques to increase the spectral efficiency and reach higher bit rates [1], [2], [3]. Indeed, it has been shown that up to 400Gb/s optical coherent transmission can be done by combining a real Analog to Digital Converter (ADC) with offline signal processing [4], [5]. Nevertheless, only 112 Gb/s coherent transmission has been experimentally tested in real-time [6], [7] and 40Gb/s (and even 100Gb/s very recently) is now proposed in commercial products [8]. Therefore coherent transmission is the leading candidate for the next generation optical transmission network at 100 Gb/s (also, called, 100Gbit Ethernet). However, due to the increase of the data traffic in a mid-term future, very high bit rate will be required (up to 1Tb/s). To satisfy such a rate, the symbol rate and the constellation size have to be increased accordingly. Unfortunately, this ultra-high data rate transmission will be more sensitive to the various signal distortions generated by the optical fiber and the transmitter/receiver devices. Consequently the main challenge will be to develop digital signal processing algorithms counteracting the propagation impairments (typically, the transmission distance is about several thousand kilometers) but compatible with electronic circuits complexity and speed.

Throughout this paper, only the linear propagation impairments listed below will be assumed. When polarization multiplexing (PolMux) is carried out, there are two kinds of interference: *i)* Inter-Symbol Interference (ISI) associated with its own polarization due to the (residual) chromatic dispersion (CD) and with the filtering effect at the transmitter and receiver sides, and *ii)* Polarization-Dependent Impairments (PDI) due to the mixing of both polarizations given rise by the polarization mode dispersion (PMD) and the polarization-dependent loss (PDL) [9], [10]. Another source of degradation concerns the phase errors which can be split into three categories: *i)* frequency offset, *ii)* constant phase offset, and *iii)* laser phase noise [11]. When the launched power is too high, some non-linear distortions such as those induced by the Kerr effect have to be taken into account as well [12], [13].

In the "signal processing" literature, numerous blind techniques[1] have been developed for mitigating the ISI/PDI, the frequency offset, and the constant phase offset. In the "optical coherent receiver design" literature, the most widespread technique for the blind ISI/PDI compensation is the Constant Modulus Algorithm (CMA) [14] and its variants such as the Radius Directed Equalizer (RDE) [15] or the Multi Modulus Algorithm (MMA) (potentially followed by the Decision-Directed (DD) algorithm) [16]. For instance, these algorithms as implemented in [16] can compensate up to 1000ps/nm of CD in a 16-QAM coherent system, and lead to reach 100 Gb/s. Notice that all the above-mentioned algorithms belong to the set of the blind linear equalizers. So far in optical communications, the blind ISI compensator has been implemented through adaptive algorithms, *i.e.*, the linear equalizer coefficients are updated as soon as one sample is incoming. Usually, for the sake of simplicity, the update equations are derived by means of the so-called stochastic gradient descent algorithm carried out either with a constant step-size (as in [14]) or with a Hessian matrix based time-varying step-size (as done in [17]).

Before going further, we remind that the propagation channel in optical communications is static over a large observation window since it varies very slowly compared to the symbol period. Indeed, the symbol period for 100 Gb/s QPSK systems is about 40 ps whereas the coherence time of the channel is of

M. Selmi, P. Ciblat, Y Jaouën and C. Gosset are with the Department of Communications and Electronics, Télécom ParisTech, France, e-mail: {mehrez.selmi,philippe.ciblat,yves.jaouen,christophe.gosset}@telecom-paristech.fr. M. Noelle is with Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany, email: markus.noelle@hhi.fraunhofer.de. This work has been funded by Institut Telecom, France in the framework of the "Future Networks Labs" and supported by the European Network of Excellence EURO-FOS.

[1]We do not consider here training approaches for which a symbol sequence known both at the transmitter and receiver sides is periodically sent in order to estimate all the impairments parameters. Then, once those parameters are estimated, impairments are mitigated using particular techniques. The description of these techniques is out of scope of this paper.

order of a few milli-seconds [18], [19], [20]. Consequently it is worth treating the data block-by-block rather than sample-by-sample. Therefore *the main contribution of this paper is to propose an implementation of all the algorithms (dealing with ISI/PDI cancellation and phase errors mitigation) in a block-wise way*. The main advantage of the block-wise approach compared to the sample-wise one is the convergence speed and the steady-state performance. Moreover, if bursty communications (with typical values of frame duration equal to a few micro-seconds) are considered, the first samples of the burst are enough to converge[2] to an adequate equalizer whereas, as we will see later, the adaptive approach has not always converged at the burst end. In addition, a lot of calculations can be also done in parallel and thus can be implemented with the current electronic devices. More precisely, *in this paper, we introduce a block-wise version of the CMA and DD equalizer in the framework of optical system architecture and an improvement of a block-wise version of a frequency offset estimator*. All the proposed algorithms work well for any PSK [3] and QAM constellation. In the simulation part, 16-QAM is considered while, in the experimental part, 8-PSK is. Notice that block-wise approaches have been already proposed for optical communications but usually either when multi-carrier transmission (such as OFDM) is employed since it is inherent to the signal structure [21], [22] or when frequency-domain equalizer (FDE) is carried out in single-carrier transmission [23], [22]. Here although working with time-domain equalizer and single-carrier transmission, we propose to estimate the transmission parameters by using a "block" manner.

One of the main drawback of the block-wise algorithms may be its less ability to track the propagation channel variation. Nevertheless, given the quasi-static property of the optical fiber channel, we will see later that our approach is still robust to its time-variation, especially to its PMD variation and also to the phase noise.

The paper is organized as follows: in Section II, the signal model, the propagation channel model and the impairments models are defined. In Section III, we introduce our block-wise blind equalizers. In Section IV, we propose a new block-wise estimator for frequency offset. In Section V, we remind some interesting results about the constant phase correction. In Section VI, we illustrate the performance of each new estimator and of the whole system via extensive simulations. In Section VII, experimental study is done. Finally concluding remarks are drawn in Section VIII.

II. Signal model

Throughout the paper, we consider only the linear impairments generated by the transmission along the optical fiber. Therefore the continuous-time received signal (in baseband) after the received filter can be written as follows

$$\mathbf{y}_a(t) = \left(\mathbf{C}_a(t) \star \mathbf{x}_a(t)\right) e^{2i\pi\delta f_a t} + \mathbf{b}_a(t) \tag{1}$$

[2]In burst mode, the algorithm is usually initialized by a trivial equalizer at each burst beginning since the CD and the PMD can be strongly different and unknown for each burst since they depend on the wavelength routing and switching.

[3]Except BPSK when CMA is carried out (for more details, see [24]).

with

- $\mathbf{y}_a(t) = [y_{a,1}(t), y_{a,2}(t)]^{\mathrm{T}}$ the bivariate received signal where $y_{a,1}(t)$ (resp. $y_{a,2}(t)$) is the received signal on X-polarization (resp. Y-polarization), and where the superscript $(.)^{\mathrm{T}}$ stands for the transposition operator.
- $\mathbf{x}_a(t) = [x_{a,1}(t), x_{a,2}(t)]^{\mathrm{T}}$ the bivariate transmitted signal where $x_{a,1}(t)$ (resp. $x_{a,2}(t)$) is the transmitted signal on X-polarization (resp. Y-polarization).
- $\mathbf{b}_a(t) = [b_{a,1}(t), b_{a,2}(t)]^{\mathrm{T}}$ the bivariate circularly-symmetric Gaussian noise with zero mean and variance N_0 per real dimension [25]. We also assume that the noise is white in time and in polarization. As it is circularly-symmetric [25], the In-phase and Quadrature components are independent and identically distributed (iid).
- the 2×2 MIMO channel whose the impulse response is given as follows

$$\mathbf{C}_a(t) = \begin{bmatrix} c_{a,1,1}(t) & c_{a,1,2}(t) \\ c_{a,2,1}(t) & c_{a,2,2}(t) \end{bmatrix}$$

  where $c_{a,p,p}(t)$ corresponds to the inter-symbol interference created by its own polarization, and where $c_{a,p,q}(t)$ ($p \neq q$) corresponds to the inter-polarization interference created by the first-order PMD phenomenon.
- δf_a is the continuous-time frequency offset expressed in Hertz.
- $\star$ stands for the convolution product.

Notice that the subscript a stands for a continuous-time/analog signal.

The transmitted signal (in baseband) on polarization p is linearly modulated by a iid sequence of QAM/PSK symbols, denoted by $\{s_p(k)\}_k$, as follows

$$x_{a,p}(t) = \sum_k s_p(k) g_a(t - kT_s) \tag{2}$$

where T_s is the symbol period and $g_a(t)$ is the shaping filter and may be, for instance, a NRZ pulse.

In order to satisfy Shannon's sampling theorem, the received signal is sampled at twice the baud rate. Due to the oversampling, no information is lost, and we can omit timing synchronization step. We thus focus on $y_p(n) = y_{a,p}(nT_s/2)$ where we remind that p stands for the polarization p. In order to "work" at the symbol rate, we stack two consecutive received samples into a bivariate process as follows

$$\mathbf{y}_p(n) = [y_{a,p}(nT_s), y_{a,p}(nT_s + T_s/2)]^{\mathrm{T}}. \tag{3}$$

Before going further let us introduce the global filter: $h_{a,p,q}(t) = c_{a,p,q}(t) \star g_a(t)$. We assume that the dispersion time of the channel is roughly upper-bounded by $(K-1)T_s$ whatever the considered polarizations. The discrete-time received signal for the polarization p takes the following form

$$\begin{aligned} \mathbf{y}_p(n) &= e^{2i\pi\tilde{\varphi}_1 n} \sum_{k=0}^{K-1} \mathbf{h}_{p,1}(k) s_1(n-k) \\ &+ e^{2i\pi\tilde{\varphi}_1 n} \sum_{k=0}^{K-1} \mathbf{h}_{p,2}(k) s_2(n-k) \\ &+ \mathbf{b}_p(n) \end{aligned} \tag{4}$$

where $\mathbf{h}_{p,q}(n) = [h_{a,p,q}(nT_s), h_{a,p,q}(nT_s+T_s/2)]^{\mathrm{T}}$, $\mathbf{b}_p(n) = [b_{a,p}(nT_s), b_{a,p}(nT_s + T_s/2)]^{\mathrm{T}}$, and $\tilde{\varphi}_1 = \delta f_a T_s/2$ is one discrete-time frequency offset. Notice, in our model, the constant phase offset is encompassed in the channel impulse response and the laser phase error is neglected. Moreover, we will assume that the channel impulse response and the frequency offset is static over the entire observation window.

The aim of the paper is to retrieve the transmitted symbols $s_p(n)$ only given the noisy observations $\mathbf{y}_p(n)$ and the signal model (Eq. (4)). To reach our goal, we will proceed into three steps as shown in Fig. 1:

- the blind ISI/PDI compensation through the evaluation of a MIMO linear fractionally spaced equalizer (FSE). By construction, our blind equalizer is robust to the presence of the frequency offset.
- the blind frequency offset (FO) estimation through the periodogram maximization. We will see that our estimator performs better if it relies on the post-equalized signal instead of on the pre-equalized signal.
- the blind constant phase estimation. After ISI/PDI and frequency compensations, the constellation may be still rotated by a constant phase since the blind equalizer has phase ambiguity. Therefore we still need to implement constant phase compensation. Adaptive version of this phase estimator will be then able to manage the presence of the laser phase noise.

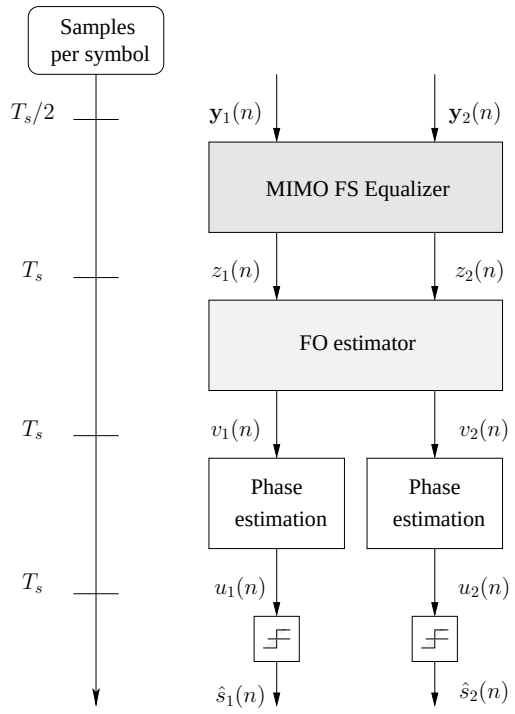

Fig. 1. Receiver structure

III. Block-wise blind equalization

In order to compensate for the channel impulse response, we introduce a $T_s/2$-fractionally spaced equalizer. Let $z_p(n)$ be the scalar output of the FSE associated with the polarization p. We have

$$z_p(n) = \sum_{k=0}^{L-1} \left(\overline{\mathbf{w}_{p,1}(k)} \mathbf{y}_1(n-k) + \overline{\mathbf{w}_{p,2}(k)} \mathbf{y}_2(n-k) \right) \quad (5)$$

where $\{\overline{\mathbf{w}_{p,q}(k)}\}_{k=0,\cdots,L}$ is the filter of length L (notice that each coefficient $\mathbf{w}_{p,q}(k)$ is a 1×2 vector, *i.e.*, corresponds to a filter with 2 inputs and 1 output) between the input polarization p and the output polarization q. The overline stands for the complex conjugation.

Eq. (5) can be re-shaped easily by means of matrices as follows

$$z_p(n) = \mathbf{w}_p^{\mathrm{H}} \mathbf{y}^{(L)}(n) \quad (6)$$

where

- $\mathbf{w}_p = [\mathbf{w}_{p,1}(0), \cdots, \mathbf{w}_{p,1}(L-1), \mathbf{w}_{p,2}(0), \cdots, \mathbf{w}_{p,2}(L-1)]^{\mathrm{T}}$,
- $\mathbf{y}^{(L)}(n) = [\mathbf{y}_1(n)^{\mathrm{T}}, \mathbf{y}_1(n-1)^{\mathrm{T}}, \cdots, \mathbf{y}_1(n-L+1)^{\mathrm{T}}, \mathbf{y}_2(n)^{\mathrm{T}}, \mathbf{y}_2(n-1)^{\mathrm{T}}, \cdots, \mathbf{y}_2(n-L+1)^{\mathrm{T}}]^{\mathrm{T}}$.
- the superscript $(.)^{\mathrm{H}}$ stands for conjugate transposition.

Notice that the filters $\mathbf{w}_{p,q}$ have $2L$ coefficients as the received signals have been sampled at twice the baud rate.

We now would like to exhibit the filter $\mathbf{w}_p$ enabling us to have $z_p(n)$ close to $s_p(n)$. To do that, it is relevant to use the CMA criterion defined as the minimization of the following cost function [26].

$$J_p(\mathbf{w}_p) = \mathbb{E}[J_{p,n}(\mathbf{w}_p)] \quad (7)$$

with

- $J_{p,n}(\mathbf{w}_p) = (|z_p(n)|^2 - R)^2$, and
- $R = \mathbb{E}[|s_p(n)|^4]/\mathbb{E}[|s_p(n)|^2]$.

Here start the main difference with the usual approach employed in coherent optical communications so far. Indeed, instead of implementing an adaptive version of this cost function, we decide to estimate the mathematical expectation of Eq. (7) given an observation block. Therefore we propose to minimize the following estimated cost function

$$\hat{J}_{p,N}(\mathbf{w}_p) = \frac{1}{N} \sum_{n=0}^{N-1} J_{p,n}(\mathbf{w}_p) \quad (8)$$

where N is the number of available quadrivariate samples $[\mathbf{y}_1(n)^{\mathrm{T}}, \mathbf{y}_2(n)^{\mathrm{T}}]$. Our purpose boils down to find the minimum of $\mathbf{w}_p \mapsto \hat{J}_{p,N}(\mathbf{w}_p)$. To do that, we suggest to use the (non-stochastic) gradient descent algorithm with optimal step size. If $\mathbf{w}_p^\ell$ is the estimated equalizer at the ℓ-th iteration (note that the data block is the same for each iteration), we have the following update relation [27][28]

$$\mathbf{w}_p^{\ell+1} = \mathbf{w}_p^\ell - \mu^\ell \boldsymbol{\Delta}^\ell \quad (9)$$

with

$$\boldsymbol{\Delta}^\ell = \left. \frac{\partial \hat{J}_{p,N}(\mathbf{w})}{\partial \overline{\mathbf{w}}} \right|_{\mathbf{w}_p^\ell}.$$

One can check that

$$\boldsymbol{\Delta}^\ell = \frac{1}{N} \sum_{n=0}^{N-1} (|z_p(n)|^2 - R) \overline{z_p(n)} \mathbf{y}^{(L)}(n) \quad (10)$$

where $z_p(n)$ is calculated by inserting $\mathbf{w}_p^\ell$ into Eq. (6).

In order to find the optimal step size μ^ℓ at the ℓ-th iteration, we minimize the estimated cost function with respect to μ^ℓ, i.e.,

$$\mu^\ell = \arg\min_{\mu} \hat{J}_{p,N}(\mathbf{w}_p^\ell - \mu\boldsymbol{\Delta}^\ell). \tag{11}$$

The derivative of $\mu \mapsto \hat{J}_{p,N}(\mathbf{w}_p^\ell - \mu\boldsymbol{\Delta}^\ell)$ is the following third-order polynomial function [27][28]

$$P^\ell(\mu) = p_3^\ell\mu^3 + p_2^\ell\mu^2 + p_1^\ell\mu + p_0^\ell \tag{12}$$

where

$$\begin{aligned} p_3^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} a_n^2, \\ p_2^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} a_n b_n, \\ p_1^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} (2a_n b_n + b_n^2), \\ p_0^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} b_n c_n \end{aligned}$$

with $a_n = |z_p(n)|^2$, $b_n = -2\Re(z_p(n)\overline{\delta}_n^\ell)$, $c_n = (|z_p(n)|^2 - R)$ and $\delta_n^\ell = (\boldsymbol{\Delta}^\ell)^{\mathrm{H}}\mathbf{y}^{(L)}(n)$.

Thanks to Eq. (12), we obtain in closed-form the roots of polynomial $P^\ell(.)$ and the real-valued root providing the minimum value of $\mu \mapsto \hat{J}_{p,N}(\mathbf{w}_p^\ell - \mu\boldsymbol{\Delta}^\ell)$ will be the selected step size at the ℓ-th iteration. Finally, the architecture of the proposed blockwise equalizer is summarized in Fig. 2.

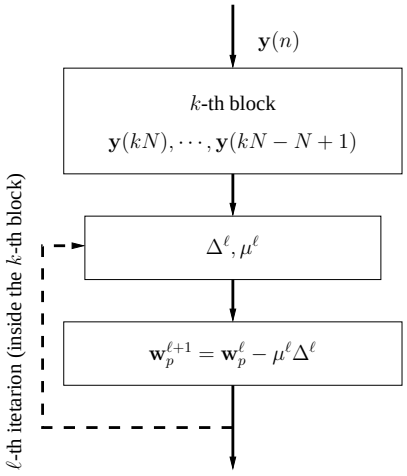

Fig. 2. Structure of the proposed block-wise equalizer

Obviously the same derivations have to be done for the polarization q. Here we decide arbitrary to treat the ISI/PDI compensation on $z_p(n)$ and $z_q(n)$ separately which implies the minimization of the following both cost functions J_p and J_q. An alternative way is possible by minimizing the mixed function $J_p + J_q$. After extensive simulations, we have remarked that such an approach leads to similar results and thus is omitted hereafter. The block-wise approach has been introduced here by minimizing the CMA criterion. It is clear that this block-wise approach can be mimicked for other criteria, such as, the Decision-Directed (DD). For example, the DD is very useful when the blind compensation has converged in order either to track slight channel modification and to improve the estimation quality.

For instance, the block-wise DD equalizer carried out with the (non-stochastic) gradient algorithm using optimal step size is very simple to implement since we are able to exhibit closed-form expression for the optimal step size. Indeed, we have

$$\boldsymbol{\Delta}_{\mathrm{DD}}^\ell = \frac{1}{N}\sum_{n=0}^{N-1} (\overline{z_p(n) - \hat{s}_p(n)})\mathbf{y}^{(L)}(n) \tag{13}$$

where $\hat{s}_p(n)$ is the current decision on the symbol $s_p(n)$. Then minimizing the function $\mu \mapsto \hat{J}_{p,N,\mathrm{DD}}(\mathbf{w}_p^\ell - \mu\boldsymbol{\Delta}_{\mathrm{DD}}^\ell)$ leads to

$$\mu_{\mathrm{DD}}^\ell = \frac{\sum_{n=0}^{N-1} \Re\{\overline{\delta_{n,\mathrm{DD}}^\ell}(z_p(n) - \hat{s}_p(n))\}}{2\sum_{n=0}^{N-1} |\delta_{n,\mathrm{DD}}^\ell|^2} \tag{14}$$

with $\delta_{n,\mathrm{DD}}^\ell = (\boldsymbol{\Delta}_{\mathrm{DD}}^\ell)^{\mathrm{H}}\mathbf{y}^{(L)}(n)$.

We now check that the complexity of the blockwise approaches is kept to reasonable values. In Table 1, we put the number of flops (complex multiplications) required for various algorithms to reach the same BER performance (in the simulation, the target-BER was fixed to 2.10^{-3} without channel coding technique). The equalizer length is fixed to $L = 3$ (*i.e.* the equalizer has 6 taps at twice baud rate). The considered algorithms are listed below:

- A-CMA: the standard Adaptive CMA.
- AN-CMA: the Adaptive CMA with Newton principle based step-size [17].
- BF-CMA: the block-wise CMA with fixed step-size (here, the step-size is $\mu = 0.02$ which is a standard value).
- BO-CMA: optimal step-size block-wise CMA.

As the number of real multiplications, divisions, additions and subtractions are negligible, and as the extraction of the third-order degree polynomial roots for the BO-CMA is also negligible, we have neglected these operations in the calculation of the computational load. Moreover, the optimal step-size associated with $J_1(.)$ can be taken almost identical to the optimal step-size associated with $J_2(.)$. Consequently, we only compute the polynomial once per iteration (either on $J_1(.)$ or on $J_2(.)$). If assuming a frame of 10000 symbols, we have a complexity of 52flops/symbol for the A-CMA and of 230flops/symbol for the BO-CMA (by considering that the equalizer obtained with the BO-CMA during the first block of length 1000 of the frame is applied on the remainder of the frame).

We remark that the BF-CMA (resp. BO-CMA) is only third times (resp. fifth times) more complex than the A-CMA but uses a much smaller set of samples. At the expense of a higher (but not unreasonable) complexity, the block-wise approaches thus converge with few samples and are especially well-adapted for burst mode transmission. Moreover the block-wise approaches are less complex than the Newton-based step-size adaptive CMA and converge much faster.

Approach	**Adaptive**		**Block** ($N = 1000$)	
Algorithm	A-CMA [14]	AN-CMA [17]	BF-CMA	BO-CMA
Update equation cost (per iteration and polarization)	$2(4L+1)$	$80L^2+8L+5$	$2N(4L+1)$	$2N(4L+1)$
Polynomial evaluation cost (per iteration)	-	-	-	$4N(3L+1)+4L$
# iterations	10000	6000	40	25
Total Flops ($\times 10^3$)	**520**	**8988**	**2080**	**2300**

TABLE I
COMPLEXITY FOR VARIOUS CMA.

IV. BLOCK-WISE FREQUENCY OFFSET ESTIMATION

Thanks to the previous section, we can now assume that ISI/PDI perfectly removed, *i.e.*, the (residual) CD and the PMD can be omitted. Therefore the (baud-rate) output of the equalizer on polarization p, already denoted by $z_p(n)$, can be written as follows

$$z_p(n) = s_p(n)e^{2i\pi(\varphi_{0,p}+n\varphi_1)} + b'_p(n) \tag{15}$$

where it remains two drawbacks:

- $\varphi_1 = \delta f_a T_s$ is the discrete-time (baud-rate) frequency offset. Notice that the frequency offset is independent of the polarization state of the received PolMux signals.
- $\varphi_{0,p}$ corresponds to the constant phase. This constant phase occurs since the blind equalizer is only able to determine the filter up to a constant phase.

and where $b'_p(n)$ is the additive zero-mean circularly-symmetric complex-valued Gaussian noise.

The construction of relevant block-wise blind estimators for the frequency offset in the context of either PSK or QAM modulations can be done by using the unique framework of the non-circularity [29], [30]. Indeed, due to rotation symmetry, it is well-known that for M-PSK, the term $\mathbb{E}[s_p(n)^Q] \neq 0$ with $Q = M$. For M-QAM, we have $\mathbb{E}[s_p(n)^Q] \neq 0$ with $Q = 4$. Then one can write $z_p(n)^Q$ as $\mathbb{E}[z_p(n)^Q] + e_p(n)$ where $e_p(n)$ is a zero-mean process that can be viewed as disturbing noise. Moreover as the noise $b'_p(n)$ is a circularly-symmetric Gaussian noise, we have that

$$\mathbb{E}[z_p(n)^Q] = \mathbb{E}[s_p(n)^Q]e^{2i\pi Q\varphi_1}.$$

Consequently, we get

$$z_p^Q(n) = A_p e^{2i\pi Q(\varphi_{0,p}+n\varphi_1)} + e_p(n) \tag{16}$$

where $A_p = E[s_p^Q(k)] \neq 0$ is a constant amplitude. The most important thing now is to remark that $z_p^Q(n)$ is actually a constant-amplitude complex exponential with frequency $Q\varphi_1$ disturbed by a zero-mean additive noise. One can thus deduce the following frequency offset estimator based on the maximization of the periodogram of $z_p^Q(n)$.

$$\hat{\varphi}_{1,N} = \frac{1}{Q}\arg\max_{\varphi\in[-\frac{1}{2},\frac{1}{2})} (f_1(\varphi) + f_2(\varphi)) \tag{17}$$

where

$$f_p(\varphi) = \left|\frac{1}{N}\sum_{n=0}^{N-1} z_p(n)^Q e^{-2i\pi\varphi n}\right|^2 \tag{18}$$

with N the number of available samples.

When PSK is encountered, our algorithm is a natural extension of the so-called Viterbi-Viterbi algorithm [31], [32] by combining linearly the periodogram obtained on each polarization. When QAM is encountered, our algorithm is also a natural extension of an existing algorithm [30]. Notice that even if the same framework enables us to treat PSK and QAM together, the performance of these algorithms are constellation-dependent. Actually, PSK works better since $\mathbb{E}[s_p(n)^Q] = s_p^Q(n)$ whereas, for QAM, $\mathbb{E}[s_p(n)^Q] \neq s_p^Q(n)$ [33].

The main issue now concerns the evaluation of the maximum in Eq. (17). Actually, in the "optical communications" literature, the maximization is done through the computation of a discrete-frequency spectrum (FFT). This FFT either has N points or has been zero-padded with αN points ($\alpha > 1$ is fixed once). Thanks to [33], the Mean Square Error (MSE) on the frequency offset decreases as $1/N^2$ for such algorithms implementation. As M-QAM is more sensitive to frequency offset, such MSE decreasing trend is not enough and more accurate estimator is required. Therefore we here propose to maximize the periodogram in different way. We compute the maximization of periodogram into two steps as follows

1) a coarse step which detects the maximum magnitude peak which should be located around the true frequency offset. This is carried out via a Fast Fourier Transform (FFT) of size N (N-FFT).
2) a fine step which inspects the cost function around the peak detected by the coarse step. This step may be implemented by a gradient-descent algorithm or the Newton algorithm [33].

Since [33], we know that the MSE associated with the algorithm carrying out the two steps decreases as $1/N^3$ and thus is significantly more accurate than the FFT based maximization.

In the second step, a Newton based gradient-descent algorithm is used, and the update equation is as follows

$$\varphi_1^{\ell+1} = \varphi_1^{\ell} + \mu\frac{f_1'(\varphi_1^{\ell}) + f_2'(\varphi_1^{\ell})}{|f_1''(\varphi_1^{\ell}) + f_2''(\varphi_1^{\ell})|}$$

with $f_p'(\varphi) = \partial f_p(\varphi)/\partial\varphi$ and $f_p''(\varphi) = \partial^2 f_p(\varphi)/\partial\varphi^2$.

As a conclusion, while the two-steps based maximization has been already used in "wireless communications" literature, the proposition of combining two periodograms of both ways (*i.e.* two polarizations in optical context or two antennas in wireless context) is new.

V. CONSTANT PHASE ESTIMATION

Thanks to previous sections, one can now assume that there is no more ISI/PDI and even no more frequency offset.

Therefore, the signal may be written as follows

$$v_p(n) = s_p(n)e^{2i\pi\varphi_{0,p}} + b''_p(n) \tag{19}$$

where $b''_p(n)$ is still an additive zero-mean circularly-symmetric complex-valued Gaussian noise.

The phase can be estimated in blindly manner via the block-wise Viterbi-Viterbi algorithm (for PSK) [31], [34] and Fourth-Power algorithm (for QAM) [30]. Thus, we have

$$\hat{\varphi}_{0,p,N} = \frac{1}{Q}\angle\left(\frac{1}{N}\sum_{n=0}^{N-1} v_p(n)^Q\right) \tag{20}$$

where $\angle(.)$ stands for the angle of complex-valued number. Notice that this previous algorithm can be improved, if necessary, by applying another (but more complicate) non-linear function to $v_p(n)$ depending on the OSNR value [29].

Once the blind phase estimator has worked, one can move to Decision-Directed phase estimator which is given by [32]

$$\hat{\varphi}_{0,p,N,\mathrm{DD}} = \angle\left(\frac{1}{N}\sum_{n=0}^{N-1} v_p(n)\overline{\hat{s}_p(n)}\right). \tag{21}$$

These estimators are already widely used by the "optical communications" community (see [14] and references therein).

In order to track the phase variation due to laser phase noise, adaptive versions of these block-wise algorithms described in Eqs (20)-(21) have to be implemented

$$\hat{\varphi}_{0,p,n+1} = \hat{\varphi}_{0,p,n} + \mu_{\mathrm{blind}}\Im[v_p(n)^Q e^{-2i\pi Q\hat{\varphi}_{0,p,n}}] \tag{22}$$

where μ_{blind} is the step size, and

$$\hat{\varphi}_{0,p,n+1,\mathrm{DD}} = \hat{\varphi}_{0,p,n,\mathrm{DD}} + \mu_{\mathrm{DD}}\Im[v_p(n)\overline{\hat{s}_p(n)}e^{-2i\pi\hat{\varphi}_{0,p,n,\mathrm{DD}}}] \tag{23}$$

where μ_{DD} is the step-size parameter as done in [16], [35].

VI. Simulation results and discussion

This section is organized as follows: in Section VI-A, we introduce the simulation set-up and especially the fiber model. In Section VI-B, we focus on the block CMA equalizer performance when the channel is either static or time-varying. In Section VI-C, we inspect the performance of the proposed frequency offset estimator.

A. Simulation set-up

Our simulation setup of the optical coherent system is as follows: a 112Gbit/s transmission is achieved by multiplexing both polarizations with 16-QAM modulated signals which corresponds to 14Gbaud transmission per polarization (which leads to $T_s = 71$ps). The transmit shaping filter is a square root raised cosine filter with a roll-off factor equal to 1. This filter is used to reduce the bandwidth of the QAM pulse since rectangular pulses produce very large frequency spectrum. The ASE noise is loaded at the receiver before a 50GHz optical filter. A matched filter associated with the shaping filter is applied at the receiver side. The continuous received electrical signal is sampled at a rate of 2 samples per symbol. A fifth-order Bessel low-pass filter with a 3dB bandwidth equal to 80% of the symbol rate was used as anti-aliasing filter.

In this section, we only simulate the main linear channel impairments in fiber-optic transmission: CD and PMD. Let $\tilde{\mathbf{C}}(\omega) = \int \mathbf{C}_a(t)e^{i\omega t}\mathrm{d}t$ be the Fourier transform of the continuous-time channel impulse response. We have

$$\tilde{\mathbf{C}}(\omega) = \tilde{\mathbf{C}}_{\mathrm{CD}}(\omega)\tilde{\mathbf{C}}_{\mathrm{PMD}}(\omega)$$

where

- the frequency channel response for the **CD** phenomenon is given by

$$\tilde{\mathbf{C}}_{\mathrm{CD}}(\omega) = \begin{bmatrix} e^{i\frac{\lambda^2\omega^2 DL_f}{4\pi c}} & 0 \\ 0 & e^{i\frac{\lambda^2\omega^2 DL_f}{4\pi c}} \end{bmatrix}. \tag{24}$$

  with the fiber length L_f, the wavelength λ, the dispersion parameter D at λ, and the light velocity c.
- the frequency channel response for the **PMD** phenomenon is given [36], [37]

$$\tilde{\mathbf{C}}_{\mathrm{PMD}}(\omega) = \mathbf{R}_\theta \mathbf{D}_{\tau_{\mathrm{DGD}},\phi}(\omega)\mathbf{R}_\theta^{-1} \tag{25}$$

  with the following birefringence diagonal matrix

$$\mathbf{D}_{\tau_{\mathrm{DGD}},\phi}(\omega) = \begin{bmatrix} e^{i(\omega\frac{\tau_{\mathrm{DGD}}}{2}+\phi)} & 0 \\ 0 & e^{-i(\omega\frac{\tau_{\mathrm{DGD}}}{2}+\phi)} \end{bmatrix} \tag{26}$$

  associated with the differential group delay between the principal states of polarizations (PSP) τ_{DGD}. Moreover, we have

$$\mathbf{R}_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{27}$$

  which represents the rotation of the reference polarization axis of the fiber's PSPs.

As the singularity issue is out of the scope of this paper, we decide to put $\phi = 0$. Indeed, such parameter choices enable us to avoid the singularity issue. Notice that there exists a few algorithms to handle the singularity issue in the literature [38], [39] which can be slightly modified in order to once again implement them in a block-wise manner.

Finally no phase noise was considered throughout this section devoted to simulation except in Fig. 13.

B. Block equalization performance

In this subsection, we firstly focus on the static channel impulse response along the entire observation window in the next paragraph. The channel is simulated as described in Section VI-A.

1) Static channel case: Except otherwise stated, in order to evaluate the performance of our algorithms, we considered the following transmission channel: the chromatic dispersion $DL_f = 1000$ps/nm (such a DL_f value corresponds to a standard residual CD), the DGD delay $\tau_{\mathrm{DGD}} = 50$ps, and the polarization rotation $\theta = \pi/4$. The OSNR (in 0.1nm) is set to 20dB. The equalizer length is fixed to $L = 3$.

We test our block-wise CMA algorithms by initializing each equalizer filter $\mathbf{w}_1$ and $\mathbf{w}_2$ with the filter $\mathbf{w}_0$ whose coefficients are 0 except the central one equal to 1. These equalizer filters are initialized with $\mathbf{w}_0$. Then, inside each block, the coefficients of these equalizer filters are updated according

to Eq. (9). When we stop to update the filter, we apply the obtained equalizer filter to the entire considered block. The BER point of any figure is obtained by averaging 150 block trials.

In Fig. 3, we depict the BER of the BO-CMA versus the number of iterations for various block sizes N. The algorithm convergence is mostly obtained for a number of iterations larger than 25. We are able to obtain a BER equal to 10^{-3} (so just below the FEC limit) when the block sizes are larger than 1000. We obviously remark that the steady state of the BO-CMA is better for large block sizes.

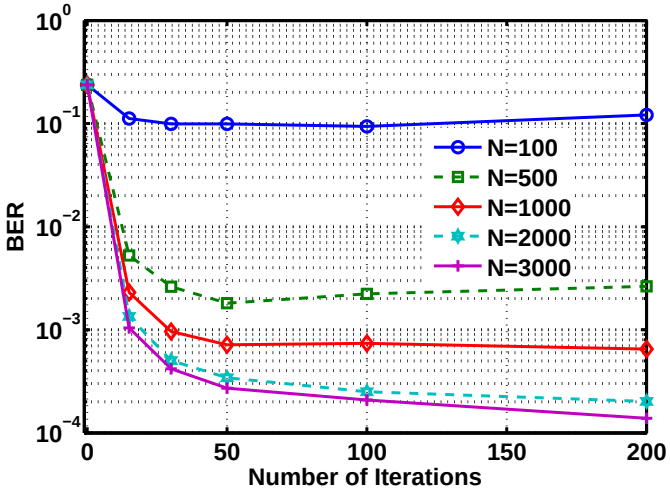

Fig. 3. BER of the BO-CMA versus the number of iterations for various block sizes N (OSNR=20dB, $DL_f = 1000$ps/nm, $\tau_{\rm DGD} = 50$ps, $\theta = \pi/4$).

In Fig. 4, we then compare the convergence speed for the BO-CMA and the BF-CMA versus the number of iterations when $N = 1000$. The BO-CMA is the fastest one since only 25 iterations were required to obtain a BER equal to 10^{-3} whereas 40 iterations are needed for the BF-CMA. However their steady-states are similar.

For the sake of clarity and simplicity, we now only display performance associated with the BO-CMA. So far, we only compare block-wise CMA algorithms to each others. To inspect the real usefulness of block-wise CMA algorithms, we will compare them (actually, only the BO-CMA) to the well-known adaptive CMA (A-CMA). In Fig. 5, we plot the BER of the BO-CMA (with 50 iterations inside each block) and the A-CMA (with fixed step-size equal to 10^{-3}) versus the observation window length. Notice that, for the BO-CMA, the observation window length is identical to the block size N, whereas, for the A-CMA, the observation window length is identical to the number of iterations. Both algorithms are initialized with $\mathbf{w}_0$ at the beginning of the observation window. We show that the BO-CMA significantly improves the convergence speed since only 1000 symbols are necessary to reach the usual target BER (around 10^{-3}) instead of 10000 for the A-CMA. Notice that the values used for Table I have been chosen according to this Fig. 5.

Until now, we only looked the performance for one block transmission. Such an approach is of interest when we would

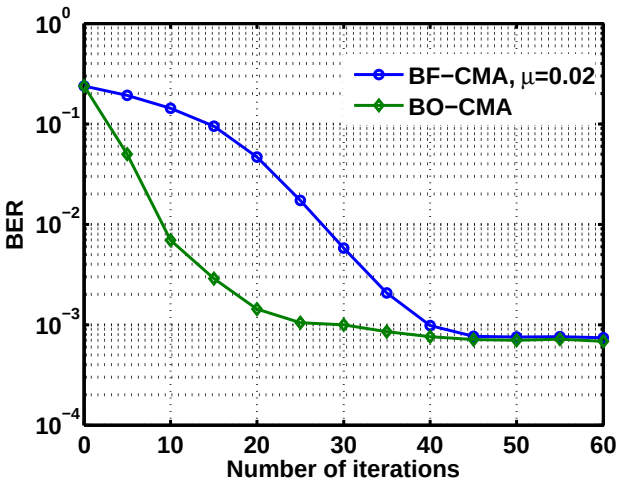

Fig. 4. BER of the BO-CMA and the BF-CMA versus the number of iterations ($N = 1000$, OSNR=20dB, $DL_f = 1000$ps/nm, $\tau_{\rm DGD} = 50$ps, $\theta = \pi/4$).

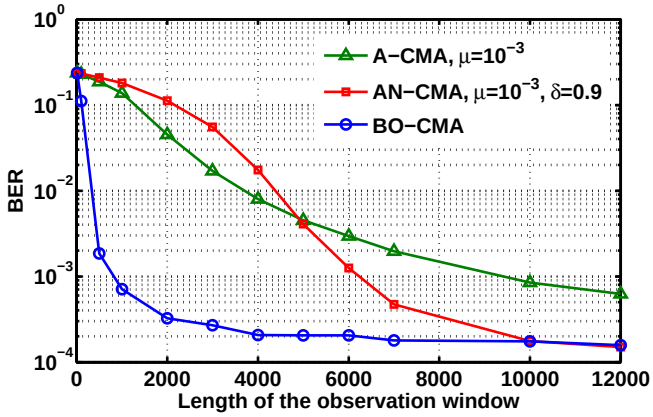

Fig. 5. BER of the BO-CMA and the A-CMA versus the observation window length (OSNR=20dB, $DL_f = 1000$ps/nm, $\tau_{\rm DGD} = 50$ps, $\theta = \pi/4$). For the BO-CMA, the observation window length is identical to the block size N. For A-CMA and AN-CMA, the observation window length is identical to the number of iterations.

like to analyze a transmission start. In the context of successive block transmission, it is clear that we have to look at the behavior of these algorithms when the initialization of the k-th block is provided by the equalizer filters obtained in the $(k-1)$-th block. Hereafter, the channel realization is still assumed to be the same whatever the considered block.

In Fig. 6, we plot the number of iterations versus the position of the block within the transmission flow. As the channel is static, we see that the number of iterations decreases with respect to the block number. It makes sense since at the beginning of the transmission (corresponding to a transition phase), the algorithm has to learn more about the channel compared to the middle and to the end of the transmission. At the end of the transmission, the algorithm is already well-initialized and just has to update slightly the equalizer coefficients. So, the more block number is high, the less

iteration number is needed. For a block size $N = 1000$, less than 10 iterations is necessary after the transition phase.

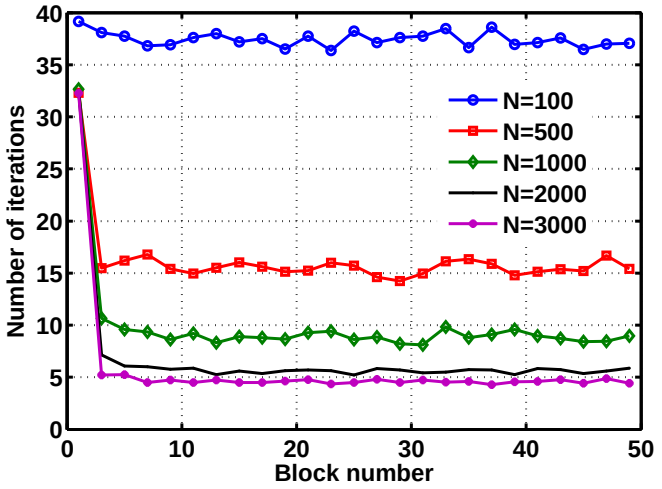

Fig. 6. The number of iterations versus the block number for the BO-CMA, (OSNR=20dB, $DL_f = 1000$ps/nm, $\tau_{\mathrm{DGD}} = 50$ps, $\theta = \pi/4$) when the proposed stopping condition is applied on the BO-CMA.

As the number of iterations depends on the block number, on the channel realization, it is worth developing a stopping criterion different from the number of iterations. We propose to stop the update when the term

$$\alpha_p^\ell = \frac{||\mathbf{w}_p^{\ell+1} - \mathbf{w}_p^\ell||}{||\mathbf{w}_p^\ell||} \tag{28}$$

is below a certain threshold. It is clear that if the steady-state is almost reached, the term α_p^ℓ will be very small. After extensive simulations not reported in this paper, we found that a target BER of 10^{-3} is usually reached when α_p^ℓ is around 5.10^{-3}. Therefore, concerning the BO-CMA, we fix the threshold for α_p^ℓ to 5.10^{-3}. To be sure to stop the algorithm (even if it does not converge) , we add a second constraint by fixing the maximum number of iterations to be equal to 40.

In Fig. 7, we plot the BER for the BO-CMA versus the block size N when the BO-CMA applied on the k-th block is initialized by the equalizer filters provided by the $(k-1)$-th block and when the aforementioned stopping condition is considered. We have observed that when the block size is too small (*e.g.*, $N = 100$), the performance are poor. The reason is that the necessary number of iterations is then higher than 40. As soon as N is large enough, the stopping condition is well-designed and the performance in terms of BER are really good.

2) Non-static channel case: In this paragraph, we would like to analyze the ability of the BO-CMA to track channel time-variation. For the sake of simplicity, we only consider infinite polarization rotation modelled by the Jones matrix. Consequently, the residual CD is assumed to be null, and the PMD only gives rise to one time-varying rotation. The polarization mixing is thus instantaneous and does not lead to inter-symbol interference (ISI) but just to inter-polarization interference (PDI). The channel impulse response at time t_0, denoted by $t \mapsto \mathbf{C}_{a,t_0}(t)$, can be written as [14]

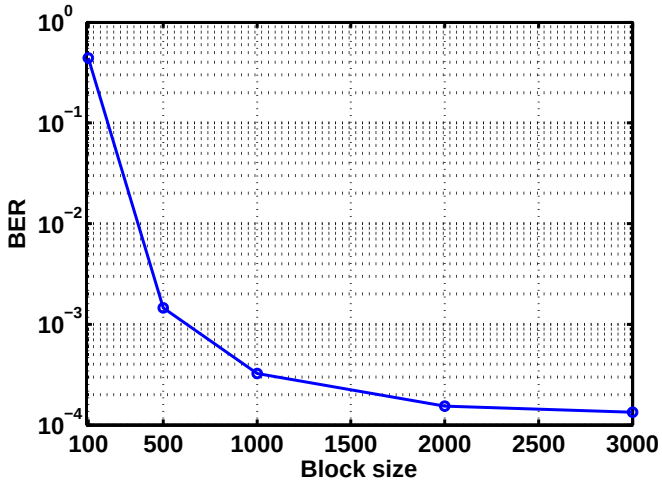

Fig. 7. BER versus the block size N for the BO-CMA (OSNR=20dB, $DL_f = 1000$ps/nm, $\tau_{\mathrm{DGD}} = 50$ps, $\theta = \pi/4$).

$$\mathbf{C}_{a,t_0}(t) = \begin{bmatrix} \cos(\Omega t_0) & \sin(\Omega t_0) \\ -\sin(\Omega t_0) & \cos(\Omega t_0) \end{bmatrix} \delta(t) \tag{29}$$

where Ω is the rotation speed in rad/s.

In Fig. 8, the BER for the BO-CMA and the A-CMA is numerically evaluated versus the rotation speed Ω. We inspect several values of the block size N in the case of the BO-CMA. We remind that the BO-CMA algorithm for the k-th block is initialized with the equalizer filters provided by the $(k-1)$-th block, and the stopping condition on α_p^ℓ is applied. The tracking ability is better for small block sizes. Moreover the BO-CMA has better tracking ability than the A-CMA. For example, for $N = 1000$, a target BER of 10^{-3} is reached up to $\Omega = 3$Mrad/s while the A-CMA is unable to track variation above $\Omega = 1$Mrad/s. The steady-state is better for larger block sizes and for low rotation speed of polarization. Notice that the steady-states are different from those offered in previous figures since the channel is built differently and is easier for small equalizer lengths due to the absence of inter-symbol interference. Besides, the smaller the block size is, the better the track ability is. As a conclusion, the block-wise CMA approach is an very promising solution since it needs smaller observation window and it offers better tracking ability.

C. Frequency offset estimation performance

In order to evaluate the performances of the proposed block-wise frequency offset (FO) estimator, we used the above described model to generate the PolMux 16-QAM signal. Except otherwise stated, we simulate a channel without CD and PMD (as explained in Section IV, we have assumed a perfect channel compensation). We then added FO randomly chosen between 0 and 3.5GHz. The FO is estimated using one of the 4 following methods:

- Coarse step based on one polarization,
- Coarse step based on both polarizations,
- Coarse and fine steps based on one polarization,

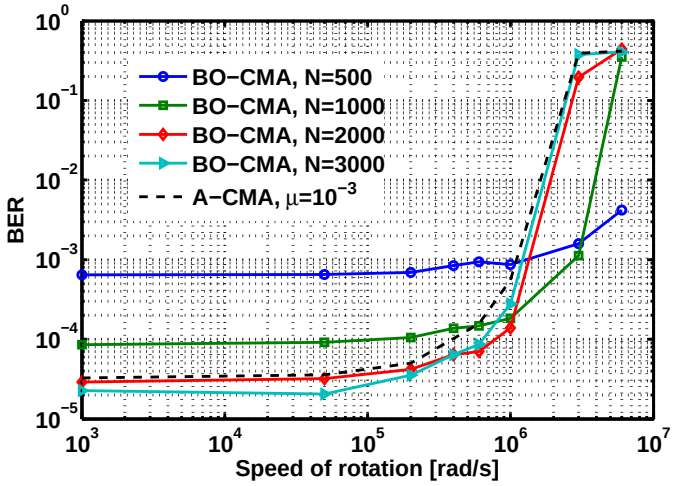

Fig. 8. BER of the BO-CMA (with different block sizes N) and the A-CMA versus the rotation speed Ω of the polarization

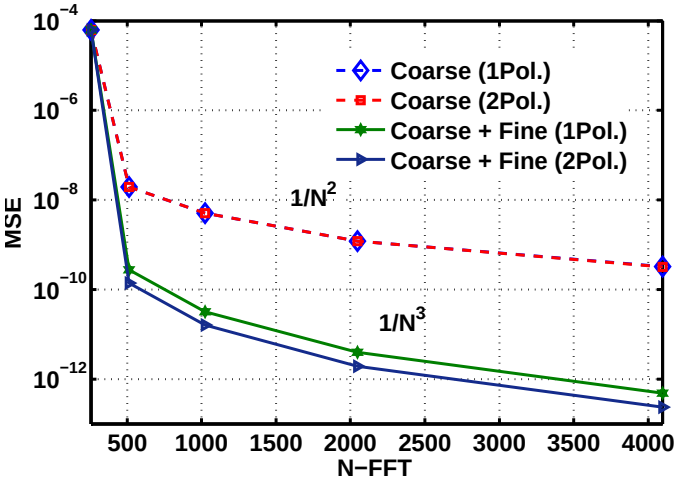

Fig. 10. MSE versus N (OSNR=18dB).

- Coarse and fine steps based on both polarizations.

The mean square error (MSE) is defined as $\mathbb{E}[|\varphi_1 - \hat{\varphi}_{1,N}|^2]$.

In Fig. 9, we plot the MSE when the FFT size (equivalently the observation window) is $N = 1024$. The most important gain in performance is due to the use of the fine step. For instance, a MSE below 10^{-11} (corresponding to some hundreds of kHz of residual FO) can be reached by using both polarizations and both steps. The outlier effect observed for low OSNR is reduced thanks to the use of both polarizations and the fine step. In Fig. 10, we plot the MSE versus N when

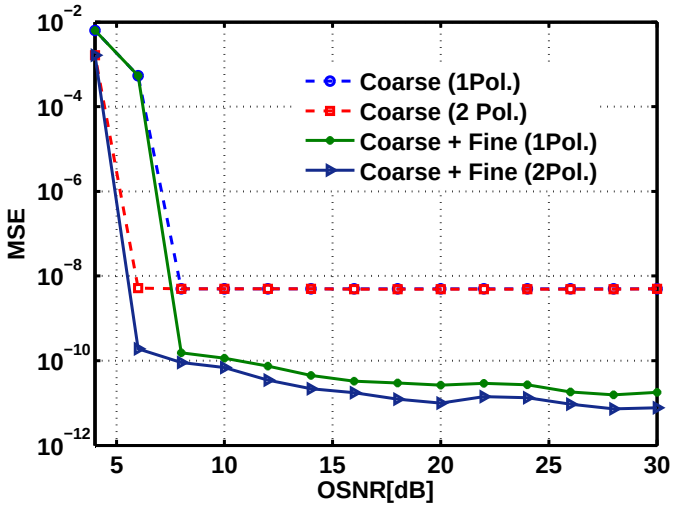

Fig. 9. MSE versus OSNR ($N = 1024$).

OSNR=18dB. One can easily check that the MSE decays as $1/N^2$ for the methods based on the coarse step and $1/N^3$ for those based on both the coarse and fine steps.

In Fig. 11, we plot the BER versus the true value of the frequency offset δf_a. The extrema for the x-label in Fig. 11 are chosen such that $4\varphi_1$ correspond to two adjacent FFT points k_0/N and $(k_0+1)/N$. Thanks to the fine step, the BER is is insensitive to the location of the true frequency offset and is below the standard target BER of 10^{-3}. In contrast, using the methods only based on the coarse step often leads to a BER much higher than the target BER.

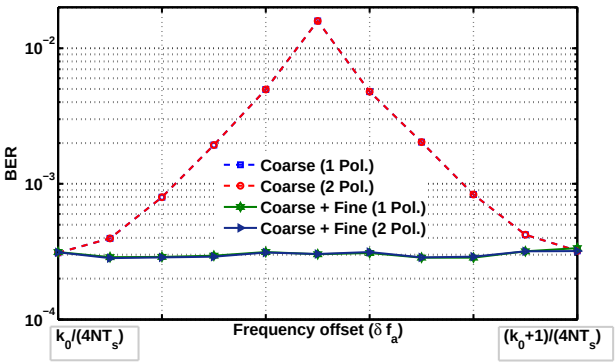

Fig. 11. BER versus the true frequency offset value for different methods ($N = 1024$, OSNR=18dB). For example, the gap in δf_a between two FFT points is here equal to 3.439MHz.

Now, we would like to analyze the robustness of the proposed FO estimator against the fiber channel impulse response and especially against the birefringence, *i.e.*, the PMD. For the sake of simplicity, we consider the following channel filter [38], [37]

$$\tilde{\mathbf{C}}_a(\omega) = \mathbf{R}_\theta \mathbf{D}_{0,\phi}(\omega).$$

We thus omit CD since $\tau_{\mathrm{DGD}} = 0$. In Fig. 12, we plot different MSE for FO estimation versus ϕ and θ defined in Eqs. (26)-(27). For each channel realization, the MSE is averaged over 100 different values of FO randomly chosen between 0 and 3.5GHz. We inspect four cases: i) one polarization based estimator implemented before CMA equalization, ii) both polarizations based estimator implemented before CMA equalization, iii) one polarization based estimator implemented after CMA equalization, iv) both polarizations based estimator implemented after CMA equalization. Notice that in order to

avoid the singularity issue, we implement the CMA proposed in [38] which handles the singularity issue [4]. When the FO estimator is implemented before PMD compensation, we fail to estimate correctly the FO when θ is between 30° and 60°. The failure probability is stronger when only one polarization is used as already seen for the outlier effect in Fig. 9. In contrast, the failure probability totally vanishes when the FO estimator is implemented after the PMD compensation. Therefore, we advocate to equalize the received signal before to estimate the FO which confirms the receiver structure described in Fig. 1.

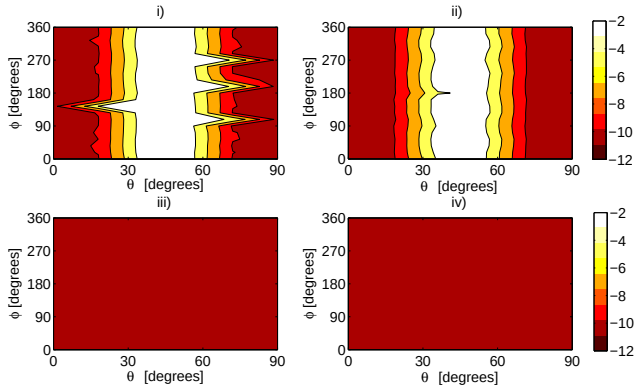

Fig. 12. $\log_{10}(\mathrm{MSE})$ ($N = 1024$, OSNR=18dB). i) one polarization before equalization, ii) both polarizations before equalization, iii) one polarization after equalization, and iv) both polarizations after equalization (when FO is chosen randomly between 0 and 3.5GHz).

Finally, we would like to inspect the impact of the phase noise on our entire receiver structure (BO-CMA, the proposed frequency offset estimator). We just proceed into two steps for handling the phase estimation in Fig. 13 . The first step consists in operating the estimator given by Eq. (20) with $N = 1000$ to counter-act the common phase. The second step will correct locally the phase noise by implementing the DD estimator given by Eq. (23). For typical value of phase noise $\Delta\nu T_s = 10^{-4}$ [16], the OSNR penalty is 1dB. Notice that more sophisticated techniques developed in [40], [41] can be also considered.

VII. Experimental results

We will validate the proposed block-based algorithms through experimental data. This enables us to investigate the effects that we did not take into account, such as, non-linear effects or non-ideal signal generation. The experimental data have been obtained by using the testbed of the *Heinrich Hertz Institute (HHI)* in Berlin. In Section VII-A, we describe the experimental set-up. In Section VII-B, the experimental performance of the BO-CMA algorithm are analyzed and compared to the A-CMA for a PolMux 8PSK transmission.

A. Experimental set-up

The experimental set-up is based on an optical 8PSK transmitter at 10GBaud corresponding to a bit rate of 30Gbit/s. The

[4]Notice that we implement the adaptive version of this CMA as in [38], and an adaptation to a block-wise version would be straightforward.

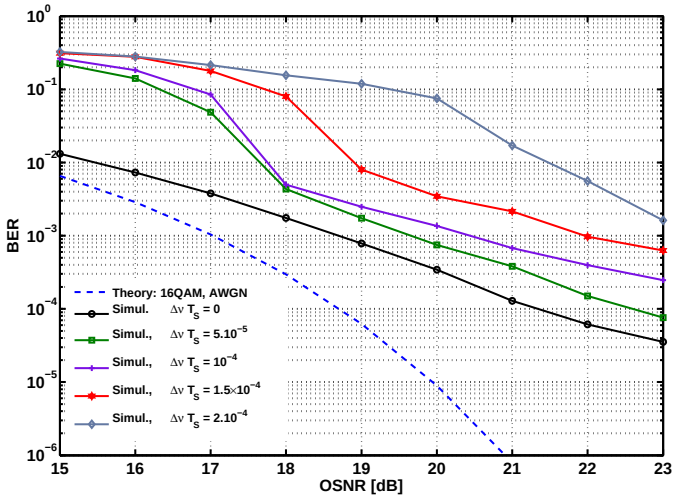

Fig. 13. BER versus OSNR ($N = 1000$, $DL_f = 1000$ps/nm, $\tau_{\mathrm{DGD}} = 50$ps, $\theta = \pi/4$). The theoretical curve corresponds to 16QAM on AWGN channel.

optical modulated signal is then multiplexed in polarization by using a polarization beam splitter (PBS) and a polarization beam combiner (PBC). A delay line is inserted into one out of the two branches in order to decorrelate the two multiplexed streams. The total bit rate of the generated PolMux 8PSK signal is thus 60Gbits/s. The transmission is performed through a recirculating loop which consists of one span of 80km of Standard Single Mode Fiber (SSMF) characterized by a cumulative dispersion of 1365ps/nm. The fiber loss is compensated for after each loop by using an Erbium-doped-fiber-amplifier (EDFA). A 5nm width filter is carried out in order to remove the out of band amplified spontaneous emission (ASE). A second EDFA is used to control the injected power at the input of each span. At the receiver side, the PolMux 8PSK signal is sent to a PBS whose outputs feed a 90^o hybrid device for each polarization. The same external cavity laser (ECL) is used for generating the 8PSK modulated signal and is shared by the local oscillator for both polarizations which implies that the frequency offset is zero [5]. The spectral linewidth of the ECL is 100kHz which leads to a no significant phase noise level. The outputs of the two 90^o hybrid devices are converted with four balanced photodiodes to generate the I and Q components for each polarization. Finally, these four signals are sampled by analog-to-digital converters at 50Gsamples/s which corresponds to 5 samples per symbol. The discrete-time data composed by 750000 samples, *i.e.*, 150000 symbols, are stored and processed offline. More details can be found in [42].

B. Performance

Except otherwise stated, we have considered a transmission over $L_f = 800$km, *i.e.*, 10 loops without inline CD compensation. The power at the input of each span was set to -0.9dBm. The cumulative CD (equal to 13650ps/nm for

[5]Consequently, the proposed frequency offset estimator is not tested with experimental data.

the 800km transmission) is partially compensated through a finite impulse response filter of length 512 [14], such as the residual cumulative CD is 1000ps/nm. On the one hand, this corresponds to practical situation when the fiber length is not perfectly known and, on the other hand, this enables us to exhibit the impact of residual CD on the proposed algorithms. The signal is then re-sampled in order to obtain exactly 2 samples per symbol. The proposed BO-CMA is finally used to compensate for the residual cumulative CD and the polarization dependent effects. As described in Section II, we compute a $T_s/2$ FSE with $L = 3$, *i.e.*, $\mathbf{w}_1$ and $\mathbf{w}_2$ have 12 complex taps each. Furthermore, we have OSNR=23.7dB.

In Fig. 14, we plot the BER versus the number of iterations inside each block for different block sizes N. The BO-CMA is initialized with $\mathbf{w}_0$ and the BER is obtained by averaging over at least 50 block observations. The target BER of 10^{-3} is obtained with a reasonable number of iterations when data blocks are larger than 500. Unlike 16QAM (see Section VI), the BO-CMA with very small block size (*i.e.*, $N = 100$) offers a higher steady-state BER. In 8PSK, the A-CMA (not plotted here) still needs tens of thousand samples to converge.

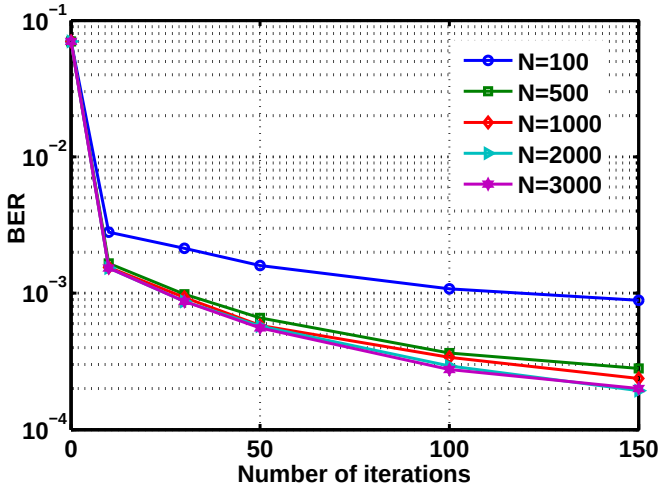

Fig. 14. BER versus the number of iterations for various N. (8PSK, OSNR=23.7dB, residual CD of 1000ps/nm, transmission distance $L_f = 800$km.)

In Fig. 15, we plot the BER versus the launched power at the input of each span for different block sizes N in order to study the influence of the intra-channel non-linear impairments. Along the data flow, the BO-CMA applied on the k-th block (of size N) is initialized with the equalizer provided by the $(k-1)$-th block (of size N). The equalizer provided by the BO-CMA (after a certain number of iterations) on the k-th block is only used on the k-th block. The number of iterations for each block is given by the stopping condition as explained in Section VI-B. Constant phase are estimated by using the algorithms described in Section V. Notice that the block size of the constant phase estimator has been fixed to 10 in order to be robust to the potential phase noise. As soon as the block size N is larger than 500, the steady-state of the BO-CMA is slightly better than that of the A-CMA (computed with $\mu = 10^{-3}$). Moreover, the BO-CMA is as robust to the non-linear effect as the A-CMA.

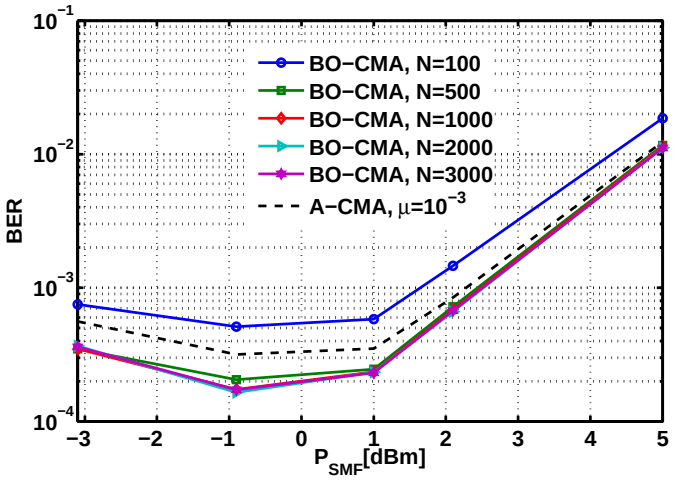

Fig. 15. BER versus the power at the input of the SSMF for various N. (8PSK, residual CD of 1000ps/nm, transmission distance $L_f = 800$km.)

In Fig. 16, we display the BER versus the residual CD. For the BO-CMA, we fix $N = 1000$. In order to handle high residual CD, the equalizer length is now increased to $L = 6$. We observe that the BO-CMA ensures slightly lower BER than the A-CMA.

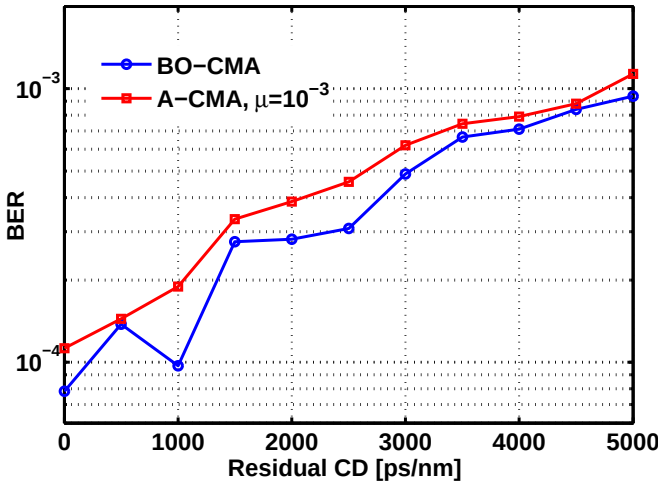

Fig. 16. BER versus the residual CD (8PSK, OSNR=23.7dB, transmission distance $L_f = 800$km.)

Even if the steady-state performance between the BO-CMA and the A-CMA are very close, we remind that the BO-CMA converges much faster than the A-CMA and thus is very well-adapted for bursty traffic mode as well as circuit mode.

VIII. Conclusion

The performance of block-wise CMA equalizer and frequency offset estimator are investigated. We showed that the observation window size required to converge for block-wise CMA approach is divided by ∼10 at the expense of an increase

of the computational complexity by a factor ~ 4. Moreover, the block-wise frequency offset estimation algorithm ensures low residual frequency offset. Finally, the block-wise digital signal processing enables us to relax the real-time implementation constraints on digital circuits running at some hundred of MHz and to offer a data throughput at a rate of tens of Gbaud. Those performances are validated through simulations and also through experimental data using a 60Gbit/s coherent optical system based on polarization multiplexing and RZ-8PSK modulation. Therefore the proposed algorithms are strong candidates for the next generation optical transmission systems.

REFERENCES

[1] X. Zhou and J. Yu, "Multi-level, multi-dimensional coding for high-speed and high-spectral-efficiency optical transmission", *IEEE Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3641-3653, Aug. 2009.

[2] M. Seimetz, "High spectral efficiency phase and quadrature amplitude modulation for optical fiber transmission - configurations, trends, and reach", Paper 8.4.3 *in Proc. of ECOC209*.

[3] M. Nakazawa, S. Okamoto, T. Omiya, K. Kasai and M. Yoshida, "256-QAM (64 Gb/s) coherent optical transmission over 160 km with an optical bandwidth of 5.4 GHz", *IEEE Photonics Technology Letters*, vol. 22, no. 3, pp. 185-187, Feb. 2010.

[4] C. Fludger *et al.*, " Coherent equalisation and POLMUX-RZ-DQPSK for robust 100-GE transmission", *IEEE Journal of Lightwave Technology*, vol. 26, no. 1, pp. 6472, Jan. 2008

[5] P. Winzer, A. H. Gnauck, S. Chandrasekhar, S. Draving, J. Evangelista, B. Zhu, "Generation and 1,200-km Transmission of 448-Gb/s ETDM 56-Gbaud PDM 16-QAM using a Single I/Q Modulator", *in Proc. ECOC'2010*.

[6] K. Fukuchi, *et al.*, "112Gb/s optical transponder with PM-QPSK and coherent detection employing parallel FPGA-based real-time digital signal processing, FEC and 100GbE Ethernet interface", Paper Tu.5.A.2 *in Proc. ECOC'2010*.

[7] C. Fludger, J.C. Geyer, T. Duthel, S. Wiese, and C. Schulien, "Real-time prototypes for digital coherent receivers", Paper OMS1 *in Proc. of OFC'2010*.

[8] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system", *Optics Express*, vol. 16, no. 2, pp. 873?879, Jan. 2008.

[9] E. M. Ip and J. M. Kahn, "Fiber impairment compensation using coherent detection and digital signal processing", *IEEE Journal of Lightwave Technology*, vol. 28, no. 4, pp 502-518, Feb. 2010.

[10] J. Renaudier *et al.*, "Linear fiber impairments mitigation of 40-Gbit/s polarization-multiplexed QPSK by digital processing in a coherent receiver", *IEEE Journal of Lightwave Technology*, vol. 26, no. 1, pp. 36-42, Jan. 2008.

[11] M. Seimetz, "Performance of Coherent Optical Square 16-QAM Systems based on IQ-Transmitters and Homodyne Receivers with Digital phase estimation ", Paper NWA4 *in Proc. OSA/NFOEC'2006*.

[12] K. Kikuchi, "Electronic post-compenation for non-linear phase fluctuations in a 1000-km 20-Gbit/s optical quadrature phase-shift keying transmission system using the digital coherent receiver", *Optics Express*, vol. 16, no. 2, pp. 889-896, Jan. 2008.

[13] J. Renaudier *et al.*, "Investigation on WDM nonlinear impairments arising from the insertion of 100-Gb/s coherent PDM-QPSK over legacy optical networks", *IEEE Journal of Lightwave Technology*, vol. 21, no. 24, pp. 1816-1818, Dec. 2009.

[14] S.J. Savory, "Digital filters for coherent optical receivers", *Optics Express*, 2005.

[15] H. Louchet, K. Kuzmin, and A. Richter, "Improved DSP algorithms for coherent 16-QAM transmission", paper Tu.1.E.6, *in Proc. ECOC'2008*.

[16] I. Fatadin, S. Savory, and D. Ives, "Blind equalization and carrier phase recovery in a 16-QAM Optical coherent system", *IEEE, Journal of Lightwave Technology*, 2009.

[17] M. Selmi, P. Ciblat, Y. Jaouen, and C. Gosset, "Pseudo-Newton based equalization algorithms for QAM coherent optical systems", *in Proc. OFC'2010*.

[18] H. Bulow, "PMD mitigation techniques and their effectiveness in installed fiber", *in Proc. OFC'2000*.

[19] E. Ip, A. Lau, D.J.F. Barros, and J.M. Kahn, "Coherent detection in optical fiber systems", *Optics Express*, vol. 16, no. 2, pp. 753-791, Feb. 2008.

[20] S. Salaun, F. Neddam, J. Poirrier, B. Rayguenes, and M. Moignard, "Fast SOP variation measurements on WDM Systems are OPMDC fast enough?", *in Proc. ECOC'2009*.

[21] J. Armstrong, "OFDM for optical communications," *IEEE Journal of Lightwave Technology*, vol. 27, no. 3, pp. 189-204, Feb. 2009.

[22] A. Barbieri et al., "OFDM vs. single-carrier transmission for 100 Gbps optical communication," *IEEE Journal of Lightwave Technology*, vol. 28, pp. 2537-2551, Sep. 2010.

[23] R. Kudo et al., "Two-stage overlap frequency domain equalization for long-haul optical systems", *in Proc. OFC'2009*.

[24] S. Houcke and A. Chevreuil, "Characterization of the undesirable global minima of Godard criterion: case of non circular symmetric complex signals", *IEEE Transactions on Signal Processing*, vol. 54, n. 5, pp. 1917-1922, May 2006.

[25] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication", Cambridge University Press, May 2005.

[26] D. N. Godard, "Self-Recovering Equalization and carrier tracking in Two-Dimensional Data Communication Systems ", *IEEE Transactions on Communications*, vol. 28, no. 11, Nov. 1980.

[27] V. Zarzoso and P. Comon, "Optimal Step-Size Constant Modulus Algorithm", *IEEE Transactions on Communications*, vol. 56, no. 1, pp. 10-13, Jan. 2008.

[28] L. Mazet, P. Ciblat and P. Loubaton, Fractionally spaced blind equalization: CMA versus second order based methods, *in Proc. SPAWC'1999*.

[29] Y. Wang, E. Serpedin, and P. Ciblat, "Optimal Blind Carrier Recovery for M-PSK Burst Transmissions", *IEEE Transactions on Communications*, vol. 51, no. 9, pp. 1571-1581, Sep. 2003.

[30] Y. Wang, E. Serpedin, and P. Ciblat, "Optimal blind nonlinear least-squares carrier phase and frequency offset estimation for general QAM modulations", *IEEE Transactions on Wireless Communications*, vol.2, no. 5, pp. 1040-1054, Sep. 2003.

[31] A.J. Viterbi and A.M. Viterbi, "Non-linear estimation of PSK-modulated carrier phase with application to burst digital transmission", *IEEE Transactions on Information Theory*, vol. 29, pp. 543-551, 1983.

[32] U. Mengali and A.N. D'Andrea, "Synchronization techniques for digital receivers", Plenum Press, 1997.

[33] P. Ciblat and M. Ghogho, "Blind NLLS Carrier frequency offset estimation for QAM, PSK, and PAM modulations: performance at low SNR ", *IEEE Transactions on Communications*, vol. 54, no. 10, Oct. 2006.

[34] D.S. Ly-Gagnon, S. Tsukamoto, K. Katoh, K. Kikuchi, "Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation", *IEEE Journal of Lightwave Technology*, vol. 24, no. 1, pp. 12-21, Jan. 2006.

[35] L. He, M.G. Amin, C. Reed, and R.C. Malkemes, "A hybrid adaptive blind equalization algorithm for QAM signals in wireless communications", *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 2058-2069, Jul. 2004.

[36] E. Ip and J.M. Kahn, "Digital Equalization of Chromatic Dispersion and Polarization Mode Dispersion", *IEEE Journal of Lightwave Technology* vol. 25, no. 8, pp 2033-2043, 2007.

[37] S.J. Savory, "Digital Coherent Optical Receivers: Algorithms and Subsystems", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp 1164-1179, Sep. 2010.

[38] X. Chongjin and S. Chandrasekhar, "Two-stage constant modulus algorithm equalizer for singularity free operation and optical performance monitoring in optical coherent receiver", *in Proc. OFC'2010*.

[39] A. Vgenis, C.S. Petrou, C.B. Papadias, I. Roudas, and L. Raptis, "Nonsingular Constant Modulus Equalizer for PDM-QPSK Coherent Optical Receivers", *IEEE Photonics Technology Letters*, vol. 22, no. 1, pp 45-47, Jan. 2010.

[40] G. Colavolpe, et al., "Robust multilevel coherent optical systems with linear processing at the receiver," *IEEE Journal of Lightwave Technology*, vol. 27, pp. 2357-2369, Jul. 2009.

[41] T. Foggi, G. Colavolpe, and G. Prati, "Stop-and-go algorithm for blind equalization in QAM single-carrier coherent optical systems," *IEEE Photonics Letters*, vol. 22, no. 24, pp. 1838-1840, Dec. 2010.

[42] M. Seimetz, L. Molle, M. Gruner, D.-D. Groß, R. Freund, "Transmission Reach Attainable for Single-Polarization and PolMux Coherent Star 16QAM Systems in Comparison to 8PSK and QPSK at 10Gbaud", Paper OTuN2 *in Proc. OFC'2009*.