Analysis of Scalable PMD Compensators using FIR Filters and Wavelength-Dependent Optical Power Measurements

Peter Oswald Christi K. Madsen Robert L. Konsbruck

***Abstract*— We investigate the potential of optical n-stage FIR filter architectures for PMD compensation based on frequency-dependent optical power monitoring. The measurement information is used to partially estimate the channel's Jones matrix $U(\omega)$, and the filter parameters are adjusted based on solving a nonlinear least-squares problem which mimics the approximate inversion of $U(\omega)$ by the filter. Numerical studies of the optimization method and outage probability simulations show the robustness and scalability of the proposed concept.**

***Index Terms*— Polarization mode dispersion compensation, optical filters, FIR filter design, numerical simulation**

I. Introduction

Polarization mode dispersion (PMD) is considered a potentially performance-limiting factor for optical communication systems operating at higher bit rates, especially, when used over legacy fiber links. A recent overview about PMD theory and modeling is given in [1]. From a practical point of view, the design and control issues related to PMD compensation (PMDC) are of particular importance for a range of applications (wideband PMD mitigation, intra-channel PMDC, higher-order PMD impairments, etc.). PMDC efforts are typically based on receiver feedback information, and may benefit from other recent signal processing and coding-theoretical advances. On the other hand, integrated all-optical PMDC solutions for use in various parts of the network are needed, too.

The existing PMDC proposals mostly neglect polarization-dependent losses and can be differentiated by a) PMDC architecture, b) monitoring information and measurement setup, c) feedback control method, and d) application range. As to architectures, most PMDC are based on combinations of polarization controllers (or equivalent combinations of phase shifters, Mach-Zehnder interferometers, waveplates, and directional couplers) and delay lines with a few tunable parameters. See [2], [3], [4], [5] for examples with up to 5 degrees of freedom. The use of n-stage finite impulse response (FIR) filter architectures was advocated in, e.g., [6], [7], [8], [9], in particular for broadband PMDC. As a basis for optical filter design, n-stage FIR filters have a well-established theory, see [10], [11], and are scalable which makes them capable of dealing with higher-order PMD. A few papers treat more general architectures by including variable delay lines (see the discussion in [3], [6], [12] and [13]) or using allpass filters based on tunable ring resonators [14].

PMD monitoring is often based on converting the optical signal into the radio-frequency domain at the receiver and directly assessing the bit error rate (BER) or eye-opening penalty, but may be too slow for use at high bit-rates, see [15] for research on this issue. As to the measurement of PMD-induced transmission degradation in the optical domain, several approaches have been pursued such as degree of polarization (DOP) [2], [16], arrival times of polarization-scrambled light [17], [4], differential group delay [18], and frequency-dependent PMD vector monitoring [19]. Only a few of the above mentioned papers give details about the potential control algorithms and their real-time behavior. Since it is generally anticipated that practical PMDCs are adjustable in a time period of a few milliseconds, and commercially available first- and second-order PMDC are only approaching this demanding requirement, much work on this issue is still ahead, especially for higher-order PMD mitigation.

The focus of this work is to further assess the potential of n-stage two-port power-complementary optical FIR filter architectures for intra-channel PMD compensation, under the assumption that a PMD monitoring scheme such as proposed in [20], [21] provides us with accurate approximations to the channel's Jones matrix $U(\omega)$ for the frequency band of interest. Based on this information, the roughly $2n$ free parameters of the PMDC which are realized in phase shifters are then adjusted such that the PMDC's Jones matrix $U_{\rm C}(\omega)$ approximately inverts $U(\omega)$, possibly up to a remainder chromatic dispersion. The optimal parameter values are found by numerically solving an associated nonlinear least-squares minimization problem, in a way that avoids sub-optimal operation. We report performance studies for the optimization and outage probability simulations that demonstrate robustness and scalability, and demonstrate the potential and current speed limitations of this approach to higher-order PMD mitigation.

Peter Oswald is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA. Email: poswald@research.bell-labs.com

Christi Madsen is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA. Email: cmadsen@lucent.com

Robert Konsbruck is with the School of Computer and Communication Sciences, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland. Email: robert.konsbruck@a3.epfl.ch. The work on this paper was done while the third author was at the Mathematics of Communications Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA. The authors would like to thank Peter Winzer, René Essiambre, and Jelena Kovačević for lending their software and generous support.

II. PMD Compensator and Monitor Architectures

The PMDC architecture considered in this paper uses a polarization beam splitter (PBS) with a polarization rotator in one arm to separate two orthogonal polarizations which we call x-polarization (depicted in Fig. 1 as the upper arm or x-arm) resp. y-polarization. The PBS is followed by an n-stage cascade of symmetrical Mach-Zehnder interferometers (sMZ), tunable phase shifters (PS) in the y-arm, and constant delay lines (DL) for the x-polarization. Depending on the application, a polarization beam combiner (PBC) is used to combine the two polarizations, or only one polarization is further transmitted. Figure 1 shows the two architectures for $n = 2$.

The constant delay ΔL is the same for all stages, and is tied to the free spectral range (FSR, in GHz) of the PMDC. As a default value, the FSR is chosen about $3/2$ times the optical filter bandwidth $B_{\rm O}$ of the receiver. In the normalized frequency variable ω (in radians), the FSR corresponds to a full period of 2π and $\Delta L = 1$. It is assumed that $\omega = 0$ corresponds to the center frequency of the channel under consideration. From now on, ω is always normalized in this fashion. The sMZ consists of a tunable PS embedded between two 50% power splitters. Thus, all tunable components are implemented as phase shifters.

To describe the transfer matrices of these components and the whole PMDC, we introduce the following basic notation. The transfer matrix of a lossless transmission link will be denoted by $T(\omega) = e^{j\phi(\omega)}U(\omega)$, where $U(\omega) \in SU(2)$ is the associated Jones matrix. The common phase factor in front of $U(\omega)$ represents the chromatic dispersion of the transmission link. Let

$$Z_{\alpha(\omega)} = \begin{pmatrix} e^{-j\alpha(\omega)/2} & 0 \\ 0 & e^{j\alpha(\omega)/2} \end{pmatrix} \tag{1}$$

denote the Jones matrix of a birefringent section with arbitrary frequency-dependent phase delay $\alpha(\omega)$. We set $c = \cos\theta$, $s = \sin\theta$, and denote by

$$R_\theta = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}, \qquad \hat{R}_\theta = \begin{pmatrix} c & -js \\ -js & c \end{pmatrix}, \tag{2}$$

the Jones matrices associated with waveplates and directional couplers (DC), respectively. Thus, a PS that realizes a phase shift by the scalar factor $e^{j\phi}$ in the y-arm has a transfer matrix $T_{\rm PS}(\omega) = e^{j\phi/2}Z_\phi$ while the transfer matrix of a DL in the x-arm is $T_{\rm DL}(\omega) = e^{-j\omega/2}Z_\omega$ (recall that $\Delta L = 1$ after normalization). Thus, the transfer matrix of a sMZ, where the phase shift by the scalar factor $e^{j\theta}$ is in the x-arm, is given by

$$T_{\rm sMZ}(\omega) = \hat{R}_{\pi/4}(e^{j\theta/2}Z_{-\theta})\hat{R}_{\pi/4} = e^{j\theta/2}R_{(\pi+\theta)/2}Z_\pi. \tag{3}$$

An obvious alternative to using a sMZ is the tunable DC, where $T_{\rm DC}(\omega) = \hat{R}_\theta$.

With this notation at hand, we see that the transfer function of an n-stage PMDC as in Fig. 1 (b) is given by

$$\begin{aligned} T_{\rm C}(\omega) \;=\; & \exp[j(\sum_{k=0}^{n}\phi_k + \sum_{k=0}^{n}\theta_k - n\omega)/2] \\ & \cdot R_{(\pi+\theta_n)/2}Z_{\omega+\phi_n+\pi}\ldots R_{(\pi+\theta_1)/2} \\ & \cdot Z_{\omega+\phi_1+\pi}R_{(\pi+\theta_0)/2}Z_{\phi_0+\pi}. \end{aligned}$$

To simplify the notation, we set $\psi_0 = (\sum_{k=0}^{n}\phi_k + \sum_{k=0}^{n}\theta_k)/2$, and replace $(\pi+\theta_k)/2$, $\phi_k+\pi$ by θ_k, ϕ_k, respectively. Then

$$T_{\rm C}(\omega) = e^{j(\psi_0 - n\omega/2)}U_{\rm C}(\omega), \tag{4}$$

where $U_{\rm C}(\omega)$ is the Jones matrix

$$U_{\rm C}(\omega) = R_{\theta_n}Z_{\omega+\phi_n}\ldots R_{\theta_1}Z_{\omega+\phi_1}R_{\theta_0}Z_{\phi_0}. \tag{5}$$

A similar representation with the R replaced by $\hat{R}$ holds if a DC replaces the sMZ. Equations (4-5) reveal that the PMDC architecture from Fig. 1 (b) is equivalent to a n-stage two-port power-complementary FIR filter (see Appendix I). The common phase factor of $T_{\rm C}(\omega)$ is given by a linear phase function which results in a constant delay of the signal, and does not impact the receiver detection.

In case of Fig. 1 (a), where a PBC combines the two polarizations before the receiver, the last sMZ and PS can be dropped from the PMDC since they correspond to a constant unitary transformation of the polarization vector, with no impact on the square-law power detection used in the receiver. Since the Jones matrix (5) of the PMDC is 2π-periodic whereas the channel's Jones matrix $U(\omega)$ is not, a good compensation result cannot be expected on a band of length close to or larger than 2π. This partly explains why we have taken the FSR sufficiently larger than the actual optical filter bandwidth B_O.

The principal structure of the PMD monitor [20], [21] is shown in Fig. 2, it can be placed before the PMDC (after the PBS) or after the PMDC (before the PBC) in Fig. 1. The co-polarized signals on the two arms are passed through a tunable narrow-band filter (NBF), and power measurements P_x, P_y are tapped off from both arms. The remainder signal is split and passed through 45° DCs (with $\theta = \pi/4$), where in one case a PS in the x-arm (with $\phi = -\pi/2$) is applied before the DC, and two more power measurements P'_x, P'_y are taken. The tuning of the NBF can be achieved by shifting the filter's center frequency across a certain set of sampling frequencies $\Omega_{\rm S}$. Altogether, for each sampling frequency $\tilde{\omega}_i \in \Omega_{\rm S}$, $i = 1, \ldots, \tilde{L}$, four power measurements will be recorded. One measurement cycle consists of a complete sweep through all $\tilde{L}$ sampling frequencies. As will be shown in Appendix II, this enables the detection of the magnitudes of the two polarizations and also the difference of their phases at a set

$$\Omega_{\rm C} := \{\omega_i,\ i = 1, \ldots, L\} \subset [-\alpha_{\rm C}\pi, \alpha_{\rm C}\pi], \tag{6}$$

of control frequencies, where $\alpha_{\rm C} < 1$. We call the interval $[-\alpha_{\rm C}\pi, \alpha_{\rm C}\pi]$ control band, for practical purposes (see Sections III and IV) it is reasonable to assume that it roughly covers the passband of the optical filter. By applying the inverse of the transfer function $T_{\rm C}(\omega)$ which is known from the

current PMDC settings, we can determine similar information about the polarization vectors $\mathbf{e}_{\mathrm{C,in}}(\omega_i)$ at PMDC input.

The accuracy of this information will depend on many factors, including the modulation format and the measurement tolerances. A first experimental validation of the above described PMD monitor is given in [21]. Measurement speeds of 3 ms for a single sweep across a modulated spectrum have been demonstrated. Preliminary DGD measurements using a periodically modulated input are in excellent agreement with a standard PMD measurement and the application to data modulated inputs and higher-order PMD is in progress. For the purpose of this paper, we will assume that the PMD monitoring process provides us with good approximations for the normalized magnitudes and phase differences of $\mathbf{e}_{\mathrm{C,in}}(\omega_i)$. This allows us to partially recover the channel's Jones matrix $U(\omega)$ at the control frequencies ω_i, and is called the *partial measurement* case. To benchmark the optimization approach and the simulations, we also have considered the *full measurement* case, where complete information about $U(\omega_i)$ is given. For a further discussion of these issues, see Appendix II and the following two sections.

III. Optimization Criteria

The monitoring of PMDC quality and parameter adjustment is based on the heuristic that under ideal circumstances the PMDC transfer matrix should undo the distortion of the channel's transfer matrix in the frequency band of interest. Roughly speaking, what we want to aim at is the *approximate inversion of* $U(\omega)$ by $U_{\mathrm{C}}(\omega)$ for $\omega \in [-\alpha_{\mathrm{C}}\pi, \alpha_{\mathrm{C}}\pi]$. Since $U_{\mathrm{C}}(\omega)$ corresponds to an arbitrary two-port power-complementary FIR filter, as the number of stages n goes to infinity we can achieve convergence

$$U_{\mathrm{C}}(\omega)U(\omega) \longrightarrow I, \qquad \omega \in [-\alpha_{\mathrm{C}}\pi, \alpha_{\mathrm{C}}\pi],$$

for any continuous Jones matrix $U(\omega)$, as long as $\alpha_{\mathrm{C}} < 1$ is fixed. The convergence speed to the limit obviously depends on the smoothness properties of $U(\omega)$ in a neighborhood of the control band; it is expected to degrade with an increase in bit rate, the amount of root-mean-square differential group delay (DGD), and higher-order PMD contents of the channel. The control band $[-\alpha_{\mathrm{C}}\pi, \alpha_{\mathrm{C}}\pi]$ is determined mainly by the optical filter bandwidth B_{O} of the receiver and the FSR of the PMDC.

Since in practice we have to fix a maximal n, we cannot hope for perfect inversion of arbitrary $U(\omega)$ by the PMDC, and need to introduce a quantitative measure of PMD distortion that can be explored in optimization/control algorithms. What we use in this paper is a modified nonlinear least-squares (NLSQ) formulation in the tuning parameters which incorporates the PMD monitor information in a straightforward way. Similar functionals have often been used for FIR filter design and in adaptive FIR filter control. Alternative formulations, which we do not pursue here, are based on weighted NLSQ resp. minmax functionals or involve the minimization of instantaneous DGD [8] resp. DOP. More precisely, given a set of unit input vectors $\tilde{\mathbf{e}}_i$ corresponding to the control frequencies $\omega_i \in \Omega_{\mathrm{C}}$, we try to determine the $2n+4$ parameters $\theta := (\theta_0, \ldots, \theta_n)$, $\phi := (\phi_0, \ldots, \phi_n)$, ψ, and τ such that

$$F(x) := \frac{1}{2}\sum_{i=1}^{L} \left\| U_{\mathrm{C}}(\omega_i)\tilde{\mathbf{e}}_i - \begin{pmatrix} e^{j(\psi+\tau\omega_i)} \\ 0 \end{pmatrix} \right\|^2 \longrightarrow \min, \tag{7}$$

where $U_{\mathrm{C}}(\omega)$ is given as a function of θ, ϕ by the formula (5), and $x := (\theta, \phi, \psi, \tau)$ is a short-hand notation for the $2n+4$ unknown variables. The input vectors $\tilde{\mathbf{e}}_i$ represent approximations to $U(\omega_i)\mathbf{e}_0$ at a fixed set of control frequencies Ω_{C}. Here $\mathbf{e}_0$ is the fixed launch polarization vector determined by the transmitter setup, and $U(\omega_i)\mathbf{e}_0$ can be interpreted as a normalization of the polarization vector $\mathbf{e}_{\mathrm{C,in}}(\omega_i)$ at the PMDC input, averaged over many bit periods. The optimization method for solving the NLSQ minimization problem (7) is briefly described in Appendix III.

Note that if

1) Ω_{C} is sufficiently dense in the control band,
2) the input vectors $\tilde{\mathbf{e}}_i$ are good approximations to $U(\omega_i)\mathbf{e}_0$, $i = 1, \ldots, L$, and
3) the value $F(x)$ of the NLSQ functional is small,

then tuning the PMDC in accordance with the parameters θ, ϕ given by x leads to a polarization vector at the PMDC output with almost all power concentrated in the x-arm, and the only additional signal degradation comes from the pre-existing chromatic dispersion (there is an additional linear phase factor, with slope depending on the parameter τ and the PMDC settings). In other words, if the Jones matrix of the system including the PMDC is denoted by $\tilde{U}(\omega) := U_C(\omega)U(\omega)$ then this should lead to an approximately constant PMD vector $\tilde{\tau}(\omega)$ of the form $\tilde{\tau}(\omega) \approx (\tilde{\tau}_0 \quad 0 \quad 0)^T$ in the control band, where $\tilde{\tau}_0$ is some real constant.

To illustrate some of the above statements, we give a couple of numerical examples (more details about the algorithm can be found in Appendix III). As input, we use

$$\tilde{\mathbf{e}}_i = U(\omega_i)\mathbf{e}_0, \qquad i = 1, \ldots, L, \tag{8}$$

where $\mathbf{e}_0$ is a fixed unit vector, where the ω_i are uniformly spaced at distance $\Delta\Omega_{\mathrm{C}} = 5$ GHz in a control band of 80 GHz length. The channel's PMD Jones matrix $U(\omega)$ is generated by a standard discrete random waveplate (DRW) model with M sections and a prescribed amount of root-mean-square (rms) DGD τ_{rmsDGD}, see the next section for details. We will characterize the impact of the PMDC optimization and its dependence on various parameters by showing the components $\tau_i(\omega)$, $i = 1, 2, 3$, of the PMD vector $\tau(\omega)$ resp. DGD values $\tau_{\mathrm{DGD}}(\omega) := \|\tau(\omega)\|$ before and after PMDC with the numerically obtained parameters. The data for the graphs is obtained by numerical differentiation from the definition of the PMD vector in terms of the corresponding Jones matrix $U(\omega)$ [1]. To make the graphs easier to compare, we have eliminated the influence of the remaining arbitrary linear phase factor after the PMDC is applied by replacing $\tau_1(\omega)$ by the deviation of $\tau_1(\omega)$ from its mean value over the control band in the definition of DGD. This modified DGD quantity is called *essential DGD* below.

To illustrate some features of our method based on the NLSQ formulation (7), we have chosen an example of a channel with relatively high DGD levels, see Fig. 3. It was generated by a DRW with $M = 200$ sections, and $\tau_{\mathrm{rmsDGD}} = 25$ ps. The actual DGD $\tau_{\mathrm{DGD}}(\omega) \approx 30$ ps of this synthetic channel is slightly higher. Fig. 4 shows the reduction of essential DGD with the number of PMDC stages n increasing. In Fig. 5, some more details about the magnitudes are shown for $n = 3$ and $n = 6$.

Note that for channel PMD content this high, we needed $n \geq 6$ stages to reduce the DGD to acceptable levels (this is not surprising as, with an FSR of 120 GHz, each stage of the PMDC can compensate for only up to ≈ 8 ps channel DGD). A further increase of n shows no visible further improvement since the number L of control frequencies and the channel PMD are fixed. Table I gives some preliminary performance measures of our Matlab implementation on a laptop with a 1 GHz Pentium III processor. We show the final value of the NLSQ functional, iteration count, and overall cpu-times (in milliseconds). Although about 200 ms per optimization run looks prohibitive compared to expected response times of a few milliseconds for practical PMDC, there is enough room for improvement since the present implementation has not been optimized towards a control algorithm, and corresponds to a cold-start situation. Also, application-dependent fine-tuning of the PMDC design parameters and tolerances used in the optimization routine might lead to further improvements (the outlier in the case $n = 8$ shows one of the rare instances, where the solver stagnates before reaching the stopping tolerance).

	Functional (7)		
n	F	Iterations	time (ms)
3	2.1193	20	230
4	1.2921	11	190
5	0.2967	19	251
6	0.0085	11	210
7	0.0030	10	220
8	0.0028	101	972

TABLE I

PERFORMANCE OF THE OPTIMIZATION METHOD: FUNCTIONAL VALUE, ITERATION COUNT, AND CPU-TIME (IN MILLISECONDS)

Finally, in Fig. 6 we illustrate the influence of the set Ω_{C} of control frequencies (c.f.). In this experiment, we have taken different numbers $6 \leq L \leq 10$ of control frequencies ω_i symmetrically located w.r.t. a control band of 80 GHz and uniformly spaced at 10 GHz distance. In all cases, we have used the same FSR of 120 GHz and $n = 6$. Obviously, for small L, there are no c.f. near the endpoints of the 80 GHz band, and the compensation quality degrades there. For $L \geq 9$, the whole control band is covered by c.f., and the DGD is uniformly reduced, at a slight PMDC degradation near the central frequency. With the exception of such boundary effects, extensive tests with various choices for Ω_{C} indicate that good quality of PMDC is achievable in a robust way with the proposed NLSQ formulation and relatively few control points. The boundary effect can be remedied in various ways, e.g., by introducing weight factors into (7) or by replacing the least-squares functional by a minmax formulation (see, e.g., [8]).

IV. OUTAGE PROBABILITY SIMULATIONS

This section reports simulation results of the performance and parameter-dependence of the proposed PMDC for signal transmission at a rate of 40 Gb/s. At the transmitter, a pseudo-random bit pattern of length 128 is launched in carrier-suppressed return-to-zero (CSRZ) modulation format with 67% duty cycle. Without loss of generality, we have assumed that all optical power is launched in the x-arm.

The channel was synthesized by a DRW model with M sections, i.e., given by a Jones matrix of the form

$$U(\omega) = \prod_{m=1}^{M} R_{-\theta_m} Z_{\tau_m \omega + \phi_m} R_{\theta_m},$$

where θ_m, ϕ_m are uniformly distributed random numbers, and τ_m is Gaussian distributed with expectation $\tau_{\mathrm{rmsDGD}}/\sqrt{M}$ and a small variance. This DRW model is standard [12] for simulating PMD effects in long lossless fiber links, and leads to a statistically correct distribution of instantaneous DGD $\tau_{\mathrm{DGD}}(\omega)$ with rms DGD given by the real parameter τ_{rmsDGD} over a wide frequency range if M is sufficiently large. We have used values of τ_{rmsDGD} between 6 ps and 20 ps, and $M \geq 50$. No other channel distortion is used in the simulations.

The receiver model takes into account optical preamplification, optical filtering (4-th order super-Gaussian filter with double-sided 3-dB bandwidth of 80 GHz), a photodiode, followed by electrical amplification, low-pass filtering (5-th order Bessel filter with single-sided 3-dB bandwidth of 28 GHz), and bit detection circuit. For the assessment of BER and outage probabilities, a quasi-analytic method based on a second order Gaussian approximation to the statistics of the photocurrent is used. This method allows us to compute the optical signal-to-noise ratio (OSNR, in dB) necessary to achieve a given BER in a deterministic way, without actually generating amplifier noise. See [22] for a more detailed description.

Below, we have fixed the BER to 10^{-9}, and recorded the OSNR penalty, i.e., the difference between the OSNR of the PMD-impaired channel (with or without PMDC) and the back-to-back OSNR, for each of 5000 random channels with rms DGD τ_{rmsDGD}. The back-to-back OSNR for the described transmitter/receiver settings was computed by feeding the transmitted signal directly into the receiver model; its level was 18.82 dB. The graphs show outage probabilities (the probability that the OSNR penalty exceeds a threshold) as functions of this OSNR threshold, and allow us to judge the impact of a given PMDC architecture, both in absolute and relative terms (when comparing with the uncompensated situation). Given that the overall ensemble consists of only 5000 channels, results for outage probabilities below 10^{-3} should be interpreted with caution.

It is clear from the previous sections that PMDC simulation results critically depend on the parameters of the PMD chan-

nel, the PMDC architecture, and the optimization framework. Generally, in what is presented below we have used a uniform 5 GHz spacing for the set of control frequencies over the control band. Stopping tolerances for the optimization routines were set to a moderate value of 0.001. The number of iterations allowed was set to 200, which in practice meant no limitations at all. If not mentioned otherwise, we have chosen an FSR of 120 GHz and a $B_{\rm C} = 80$ GHz wide control band, and dropped the second polarization before the receiver (the small power penalty associated with this will be discussed below). The approximate $3:2$ ratio between FSR and $B_{\rm C}$ as well as tying the control bandwidth to the optical filter bandwidth $B_{\rm C} \approx B_{\rm O}$ appear to be sound choices. When varying these parameters, we found only small differences, which suggest that other choices are not likely to lead to substantial improvements. Fig. 7 shows the outage probability curves for various combinations of FSR and $B_{\rm C}$ for $\tau_{\rm rmsDGD} = 10$ ps and $n = 3$. Note that reducing FSR resp. control bandwidth slightly below their $120/80$ GHz default values generally improves the PMDC result for some of the channels with higher actual DGD values, whereas a small distortion of the PMDC result appears for all channels. The latter obviously comes from the slightly increased portion of the optical filter bandwidth where the PMDC response is uncontrolled and may be arbitrary. For $B_{\rm C} < 60$ GHz, this boundary effect becomes dominating, while for $B_{\rm C} > 85$ GHz, results deteriorate in the range of outage probabilities below $< 10^{-2}$.

The set of graphs in Fig. 8 shows the effect of PMDC as a function of n and $\tau_{\rm rmsDGD}$. We show outage probability curves for rms DGD values of 6 ps, 10 ps, and 20 ps (from top to bottom), and various n. The graphs demonstrate that, in order for the outage probability to remain under 10^{-3} for a given OSNR penalty threshold of, say, 1 dB, the number of stages n should grow about linearly with $\tau_{\rm rmsDGD}$. They also show that the proposed algorithms scale well, and that more stages are needed for handling larger DGD amounts. Finally, the results seem to indicate that one- and two-stage compensators will probably reach their practical limits if the rms DGD exceeds the 10 ps level.

The impact of including a linear phase term $e^{j\tau\omega}$ into the optimization functional as well as only feeding one polarization into the receiver is illustrated in Fig. 9, again for $n = 3$ and $\tau_{\rm rmsDGD} = 10$ ps. The additional degree of freedom τ definitely helps to improve the results for the high-DGD cases. The further improvement we get by dropping the contribution of the second polarization needs some commentary. The graph in Fig. 9 does not include the penalty due to signal power reduction associated with dropping one polarization. However, a monitoring of the power loss factor, i.e., the ratio of signal power received from one polarization versus both polarizations combined via a PBC, revealed that the worst case power penalty above was 0.2 dB, whereas the average penalty amounted to 0.03 dB. Note that these numbers are similar for all cases shown in the previous graphs, with the exception of very small n, where the worst case power penalty (not the average penalty!) reached 4 dB in one outlier case. However, for these values of n, the PMDC result would not have been considered satisfactory anyway.

In Fig. 10, we show the detoriation due to various levels of noise in the power measurements. More precisely, instead of using (8) directly, we have computed input information from sets of power measurements (as explained in Appendix II) for two non-collinear input polarizations. To simulate noise, we added random Gaussian noise with variance σ to the exact power values, independently for each sampling frequency $\tilde{\omega}_i$. The plot shows the $\tau_{\rm rmsDGD} = 10$ ps, $n = 3$ case, and demonstrates sufficiently robust behavior with respect to moderate measurement errors.

For all of the above graphs, we have used (8) as input data for the optimization. These results for the full measurement case give a baseline for what could theoretically be achieved, and should be contrasted with what can be expected if only partial measurements are assumed. As will be worked out in Appendix II, with the input polarization fixed the proposed monitoring process recovers good approximations to $\mathbf{e}_{\rm C,in}(\omega_i)$ only up to a generally unknown common phase factor, and leaves us after optimization with an additional PMDC-induced chromatic dispersion. The penalty associated with this remainder dispersion is visualized in the graphs of Fig. 11. We show only the 10 ps rms DGD case.

V. Conclusions

We have described a scalable PMD compensation strategy for single channel applications based on a standard n-stage FIR filter architecture and frequency-dependent power measurements for PMD monitoring. The tunable parameters of the PMDC are found by solving a nonlinear least-squares problem whose practical solution is detailed. Parameter studies and system simulations show the feasibility of the approach, and provide hints at possible improvements (dropping of the second polarization, remainder linear phase, sampling densities, FSR/control bandwidth) but also deficiencies in the partial measurement case. Among the challenges ahead is the experimental validation and hardware integration of the PMD monitoring and PMDC tuning process, and the real-time implementation of a PMDC control algorithm based on the currently pursued optimization methods. We anticipate that a combination of improvements in several areas (fiber design, modulation formats, coding schemes, etc.) with current PMDC efforts will be necessary to provide acceptable practical solutions for rates beyond 40 Gb/s.

References

[1] H. Kogelnik, L. E. Nelson, and R. M. Jopson, "Polarization mode dispersion," in *Optical Fiber Communication Systems*, I. P. Kaminov and T. Li, eds., Acad. Press, 2002, vol. IVb, ch. 15, pp. 725–861.

[2] F. Roy, C. Francia, F. Bruyere, and D. Penninckx, "A simple dynamic polarization mode dispersion compensator," in *Optical Fiber Communications Conference*, 1999, TuS4, pp. 275–278.

[3] H. Sunnerud, C. Xie, M. Karlsson, R. Samuelsson, and P. A. Andrekson, "A comparison between different PMD compensation techniques," *J. Ligthw. Technol.*, vol. 20, pp. 368–378, Mar. 2002.

[4] D. Sandel, F. Wüst, V. Mirvoda, and R. Noé, "Standard (NRZ 1×40 Gb/s, 210 km) and polarization multiplex (CS-RZ 2×40 Gb/s, 212 km) transmissions with PMD compensation," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1181-1183, Aug. 2002.

[5] T. Saida, K. Takiguchi, S. Kuwahara, Y. Kisaka, Y. Miyamoto, Y. Hashizume, T. Shibata, and K. Okamoto, "Planar lightwave circuit polarization mode dispersion compensator," in *European Conference on Optical Communications*, 2001, Paper Mo.F.2.5, pp. 10–11.
[6] R. Noé, D. Sandel, M. Yoshida-Dierolf, S. Hinz, V. Mirvoda, A. Schöpflin, C. Glingener, E. Gottwald, C. Scheerer, G. Fischer, T. Weyrauch, and W. Haase, "Polarization mode dispersion compensation at 10, 20, and 40 Gb/s with various optical equalizers," *J. Lightw. Technol.*, vol. 17, pp. 1602–1616, Sept. 1999.
[7] L. Möller, "Filter synthesis for broad-band PMD compensation," *IEEE Photon. Technol. Lett.*, vol. 12, 1258–1260, Sept. 2000.
[8] A. Eyal and A. Yariv, "Design of broad-band PMD compensation filters", *IEEE Photon. Technol. Lett.*, vol. 14, 1088–1090, Aug. 2002.
[9] M. Bohn, G. Mohs, C. Scherer, C. Glingener, C. Wree, and W. Rosenkranz, "An adaptive optical equalizer concept for single channel distortion compensation", in *European Conference on Optical Communications*, 2001, Paper Mo.F.2.3, pp. 6–7.
[10] Z. Jinguji and M. Kawachi, "Synthesis of coherent two-port lattice-form optical delay-line circuit", *IEEE J. Lightwave Technology*, vol. 13, pp. 73–82, Jan. 1995.
[11] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: PTR Prentice Hall, 1993.
[12] Y. Li, A. Eyal, and A. Yariv, "Higher order error of discrete fiber model and asymptotic bound on multistaged PMD compensation," *J. Lightw. Technol.*, vol. 18, pp. 1205–1213, Sept. 2000.
[13] M. Shtaif, A. Mecozzi, M. Tur, and J. A. Nagel, "A compensator to the effects of high-order polarization mode dispersion in optical fibers," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 434–436, Apr. 2000.
[14] C. Madsen and P. Oswald, "Optical Filter Architecture for Approximating any 2x2 Unitary Matrix" *Optics Lett.*, vol. 28, pp. 534–536, Jul. 2003.
[15] F. Buchali, W. Baumert, H. Bülow, J. Poirrier, and S. Lanne, "A 40 Gb/s eye monitor and its application to adaptive PMD compensation", in *Optical Fiber Communications Conference*, 2002, Paper WE6, pp. 202–203.
[16] N. Kikuchi, "Analysis of signal degree of polarization degradation used as control signal for optical polarization mode dispersion," *J. Lightw. Technol.*, vol. 19, pp. 480–486, Apr. 2001.
[17] R. Noé, D. Sandel, V. Mirvoda, F. Wüst, and S. Hinz, "Polarization mode dispersion detected by arrival time measurement of polarization-scrambled light," *J. Lightwave Technol.*, vol. 20, pp. 229–235, Feb. 2002.
[18] L.-S. Yan, Q. Yu, A. B. Sahin, and A. E. Willner, "Differential group delay monitoring used as feedforward information for polarization mode dispersion compensation," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1463–1465, October 2002.
[19] L. Möller and L. Buhl, "Method for PMD vector monitoring in picosecond pulse transmission systems," *J. Lightw. Technol.*, vol. 19, pp. 1125–1129, Aug. 2001.
[20] C. Madsen, E. Laskowski, J. Bailey, E. Chen, A. Griffin, M. Capuzzo, L. Gomez, R. Long, L. Stulz, and A. Wong-Foy, "An integrated wavelength-sensitive polarimeter," in *European Conference on Optical Communications*, 2002, Paper 1.2.5.
[21] C. Madsen, P. Oswald, S. Chandrasekhar, L. Buhl, M. Capuzzo, E. Laskowski, E. Chen, L. Gomez, A. Griffin, A. Kasper, L. Stulz, and A. Wong-Foy, "Real-time PMD measurement using an integrated wavelength-scanning polarimeter," *European Conference on Optical Communications*, 2003, Paper Th2.2.5.
[22] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, and Techniques*. Kluwer Acad./Plenum Publ., 2000.
[23] "*Optimization Toolbox User's Guide*," Version 2, *The MathWorks, Inc.*, Natick MA, 2000.

APPENDIX I
FACTORIZATION OF JONES MATRICES

The PMDC filter architectures under consideration are intimately connected with well-known factorization techniques for 2×2 polynomial Jones matrices [11], [10]. Since we make use of these factorizations in the optimization scheme, we will give some details (the notation is slightly adapted to our needs). A 2×2 matrix function $V_n(z)$ will be called *(polynomial) Jones matrix of degree* n if it is of the form

$$V_n(z) := \begin{pmatrix} z^{-n/2}p(z) & z^{-n/2}q(z) \\ -z^{n/2}q(z)^* & z^{n/2}p(z)^* \end{pmatrix}, \tag{9}$$

where $p(z)=\sum_{k=0}^{n} p_k z^k$, $q(z)=\sum_{k=0}^{n} q_k z^k$ are complex polynomials that satisfy the power-complementarity (PC) condition

$$|p(z)|^2+|q(z)|^2=1, \qquad z:=e^{j\omega}, \quad -\infty<\omega<\infty. \tag{10}$$

Jones matrices of degree 0 are just the constant Jones matrices, and Jones matrices of degree 1 can alternatively be written in the form $V_1(z)=WZ_\omega V^*$, where W,V are constant Jones matrices. Note that (9) describes an arbitrary two-port PC-FIR filter since it can be rewritten in matrix-polynomial form as follows:

$$V_n(z)=z^{-n/2}\sum_{k=0}^{n} H_k z^k, \quad H_k=\begin{pmatrix} p_k & q_k \\ -q_{n-k}^* & p_{n-k}^* \end{pmatrix}, \tag{11}$$

$k=0,\dots,n$.

Any Jones matrix of degree n can be factored into n Jones matrices of degree 1 by an elementary recursive procedure (see [11]),

$$V_n(z)=V_1^{(n)}(z)\dots V_1^{(1)}(z)=U^{(n)}Z_\omega\dots U^{(1)}Z_\omega U^{(0)}, \tag{12}$$

where the $U^{(k)}$ are constant Jones matrices. For the basic factorization step, i.e., to find a factor $V_1(z)$ of $V_n(z)$ such that $V_n(z)=V_1(z)V_{n-1}(z)$, all one needs is a vector $v\neq 0$ such that $H_n v=0$ (that H_n as well as H_0 are singular follows from the PC condition (10), see [11] for details). Finally, note that any constant Jones matrix U can be further factored in various ways. Since we propose to use symmetrical Mach-Zehnder interferometers and phase shifters in our PMDC architecture, our choice of parametrization is given by

$$U=Z_{\psi_1}R_\theta Z_{\psi_2}, \tag{13}$$

which holds for any U with some ψ_i,θ. Thus, two phase shifters and one waveplate (or one Mach-Zehnder interferometer) are needed to implement an arbitrary constant Jones matrix U. Substituting (13) into (12), we see that an arbitrary Jones matrix of degree n can be written as

$$V_n(z)=Z_{\phi_{n+1}}R_{\theta_n}Z_{\omega+\phi_n}\dots R_{\theta_1}Z_{\omega+\phi_1}R_{\theta_0}Z_{\phi_0}, \tag{14}$$

i.e., parametrized by $2n+3$ parameters $\phi_0,\dots,\phi_{n+1}$, $\theta_0,\dots,\theta_n$. In practical applications, the last phase shifter represented by $Z_{\phi_{n+1}}$ can be dropped (it contributes a frequency-independent constant phase shift to each of the polarizations), and we arrive at the $(2n+2)$-parameter model (5) for the Jones matrix $U_C(\omega)$ of the n-stage FIR filter PMDC architecture adopted in this paper.

The factorization procedure provides a simple switch between the polynomial expressions in (9) and (11), which allows us to adapt the filter transfer function to the measured channel response by means of polynomial approximation, on the one hand, and the tunable parameters of the PMDC implementation, on the other. However, in practice, even if

the detected channel transfer matrix has been normalized such that it is represented by a Jones matrix, standard polynomial approximation methods will not guarantee the PC condition (10). In this situation, we propose an *approximate factorization method* based on the following heuristics. Let $V_n(z)$ be of the form (11), and assume that the PC condition holds only approximately. In this case, the matrix H_n is not necessarily singular, and may not have a null vector v. Thus, an exact factorization does not hold. However, since $|p(z)|^2+|q(z)|^2 \approx 1$ it is reasonable to assume that H_n is close to singular; in particular, that one of its singular values is close to 0. Let $H_n = \tilde{U}\Sigma\tilde{V}^*$ be the singular value decomposition of H_n, where $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2)$ is such that $0 \le \sigma_1 \le \sigma_2$. If we now set $V_1(z) = \tilde{V} Z_\omega \tilde{V}^*$ then, following the same procedure as in the case when (10) holds, we can find a Jones matrix $V_{n-1}(z)$ of degree $n-1$ such that the deviation $\|V_n(z) - V_{n-1}(z)V_1(z)\|$ is bounded by $2\sigma_1$, and thus small as long as σ_1 is small.

After this we can repeat the procedure with $V_{n-1}(z)$, and by induction we will arrive at an approximate factorization of $V_n(z)$ into n Jones matrices $V_1^{(k)}(z)$ of degree 1 such that

$$\|V_n(z) - V_1^{(n)}(z)\dots V_1^{(1)}(z)\| \le 2(\sigma_1^{(1)} + \dots + \sigma_1^{(n)}).$$

Although, due to the recursive nature of the construction, this is not an efficient error estimate (the $\sigma_1^{(k)}$ values tend to grow exponentially with k), and we do not consider this a robust procedure for larger n resp. large deviations from the PC condition, we have successfully used it for the simulations with $n \le 10$ reported in this paper. To simplify the above described algorithm further, we have avoided computing SVDs, and chosen the unit vector v such that it is orthogonal to the row of H_n with the larger Euclidean norm. This leads to qualitatively the same factorization error for $n \le 10$. For larger n, spectral factorization techniques are usually recommended [11], [10], [8], but there was no need in pursuing them for the parameter ranges presently considered.

APPENDIX II
MATHEMATICAL MODEL OF THE PMD MONITOR

Our PMDC approach assumes that at input of the optimization procedure, approximations $\tilde{\mathbf{e}}_i$ to the normalized polarization vector at PMDC input are given for a sufficiently dense set of control frequencies ω_i, $i = 1, \dots, L$. In the simulations reported on in Section IV, we have mostly addressed the full measurement case, where $\tilde{\mathbf{e}}_i$ is given by (8). Here we want to give more details on the PMD monitor introduced in Section II, and discuss how close we can come to this somewhat idealistic assumption by relying on frequency-dependent power measurements. With one fixed input polarization vector, the PMD monitoring process corresponds to the partial measurement case briefly presented in Section IV.

Throughout the derivation, we assume that the PMD monitor comes after the PMDC. Let $T(\omega) = e^{j\phi(\omega)}U_{\mathrm{C}}(\omega)U(\omega)$ denote the transfer matrix of the system, and assume that at the input to the fiber link, the polarization vector is given by $\mathbf{e}_{\mathrm{in}} := s(\omega)\mathbf{e}_0$ where the scalar function $s(\omega)$ contains the signal information, and $\mathbf{e}_0$ is a constant unit vector. Consequently,

$$\mathbf{e}_{\mathrm{out}}(\omega) = T(\omega)\mathbf{e}_{\mathrm{in}}(\omega) = \tilde{s}(\omega)U_{\mathrm{C}}(\omega)U(\omega)\mathbf{e}_0,$$

where $\tilde{s}(\omega) := s(\omega)e^{j\phi(\omega)}$. Denote the components of $\mathbf{e}_{\mathrm{out}}(\omega)$ by $x_1(\omega) := |x_1(\omega)|e^{j\phi_x(\omega)}$ and $y_1(\omega) := |y_1(\omega)|e^{j\phi_y(\omega)}$. A close look at the monitor architecture outlined in Section II reveals that as the result of the measurement process we obtain 4 vectors P_x, P_y, P'_x, P'_y the components of which contain approximate information about $\mathbf{e}_{\mathrm{out}}(\omega)$ as follows:

$$\begin{aligned} P_{x,i} &\approx |x_1(\omega_i)|^2, \qquad P_{y,i} \approx |y_1(\omega_i)|^2, \\ P'_{x,i} &\approx \frac{1}{2}|x_1(\omega_i) + y_1(\omega_i)|^2, \\ P'_{y,i} &\approx \frac{1}{2}|-jx_1(\omega_i) + y_1(\omega_i)|^2, \end{aligned}$$

where $\omega_i \in \Omega_{\mathrm{C}}$, $i = 1, \dots, L$ (we have assumed that $\Omega_{\mathrm{S}} = \Omega_{\mathrm{C}}$ for simplicity only). Since

$$\begin{aligned} |x_1(\omega_i) + y_1(\omega_i)|^2 &= |x_1(\omega_i)|^2 + |y_1(\omega_i)|^2 \\ &\quad +2\mathrm{Re}(x_1(\omega_i)y_1(\omega_i)^*), \\ |-jx_1(\omega_i) + y_1(\omega_i)|^2 &= |x_1(\omega_i)|^2 + |y_1(\omega_i)|^2 \\ &\quad +2\mathrm{Im}(x_1(\omega_i)y_1(\omega_i)^*), \end{aligned}$$

and $x_1(\omega_i)y_1(\omega_i)^* = |x_1(\omega_i)||y_1(\omega_i)|e^{j(\phi_x(\omega_i)-\phi_y(\omega_i))}$, we see from this that the PMD monitor can sample the *magnitudes* of the two components of $\mathbf{e}_{\mathrm{out}}(\omega)$ and their *phase difference* $\phi_x(\omega) - \phi_y(\omega)$ but not the individual phase values $\phi_x(\omega)$, $\phi_y(\omega)$ (in other words, we know the Stokes space representation of $\mathbf{e}_{\mathrm{out}}(\omega)$). For short, we will say that $\mathbf{e}_{\mathrm{out}}(\omega)$ is determined *up to a common phase factor*, i.e., up to a scalar factor of the form $e^{j\psi(\omega)}$ with unknown function $\psi(\omega)$ to be applied to both components of the vector.

Since we know $U_{\mathrm{C}}(\omega)$ from the current PMDC parameter settings, we can find approximate sampling information for the polarization vector

$$\mathbf{e}_{\mathrm{C,in}}(\omega) := U_{\mathrm{C}}(\omega)^*\mathbf{e}_{\mathrm{out}}(\omega) = \tilde{s}(\omega)U(\omega)\mathbf{e}_0,$$

at PMDC input, again up to a common phase factor. Assuming that the signal power (averaged over many bit periods) is not negligible in the control frequency band, we can normalize $\mathbf{e}_{\mathrm{C,in}}(\omega)$ to 1. If we denote the components of this normalized vector by $u(\omega)e^{j\phi_u(\omega)}$ and $v(\omega)e^{j\phi_v(\omega)}$, then our derivation tells us that we have samples of the magnitudes $u(\omega)$, $v(\omega)$, where $u(\omega)^2 + v(\omega)^2 = 1$ by normalization, and again of the phase difference $\Delta\phi(\omega) := \phi_u(\omega) - \phi_v(\omega)$. Thus, we can define

$$\tilde{\mathbf{e}}(\omega) := e^{j\tilde{\phi}(\omega)}(u(\omega)e^{j\Delta\phi(\omega)/2} \;\; v(\omega)e^{-j\Delta\phi(\omega)/2})^T, \quad (15)$$

where $\tilde{\phi}(\omega)$ is a phase factor at our disposal which we will fix later. Obviously, $\tilde{\mathbf{e}}(\omega)$ coincides with the normalized $\mathbf{e}_{\mathrm{C,in}}(\omega)$ up to an unknown phase factor.

From (15), we compute the input data $\tilde{\mathbf{e}}_i \approx \tilde{\mathbf{e}}(\omega_i)$ for the minimization problem (7). Suppose that the optimization terminates at a solution $x = (\theta, \phi, \psi, \tau)$ to (7) with small value of the NLSQ functional $F(x)$. Then, after tuning to the new PMDC parameters, we have a new Jones matrix $\tilde{U}_{\mathrm{C}}(\omega)$ for which $\tilde{U}_{\mathrm{C}}(\omega)\tilde{\mathbf{e}}(\omega) \approx (e^{j(\psi+\tau\omega)} \;\; 0)^T$ in the control band. Since

$\tilde{\mathbf{e}}(\omega)$ and $\mathbf{e}_{\mathrm{C,in}}(\omega)$ are the same up to the common factor $\tilde{s}(\omega)$ and another common phase factor depending on the choice we make for $\tilde{\phi}(\omega)$, we finally arrive at the following approximate expression for the compensated polarization vector at PMDC output:

$$\tilde{\mathbf{e}}_{\mathrm{out}}(\omega) := \tilde{U}_{\mathrm{C}}(\omega)\mathbf{e}_{\mathrm{C,in}}(\omega) \approx s(\omega)(e^{j(\psi+\tau\omega+\tilde{\psi}(\omega))}\ 0)^T. \tag{16}$$

The generally nonlinear phase function $\tilde{\psi}(\omega)$ contains the existing chromatic dispersion represented by the common factor $e^{j\phi(\omega)}$ in $T(\omega)$, and an additional contribution due to the incomplete phase detection resp. the choice of $\tilde{\phi}(\omega)$ in (15). We call the latter *PMDC-induced chromatic dispersion*. Currently we determine the unknown $\tilde{\phi}(\omega)$ such that the input vectors $\tilde{\mathbf{e}}_i \approx \tilde{\mathbf{e}}(\omega_i)$, while having the prescribed magnitudes and phase differences for their entries, change as smoothly as possible between neighboring sampling frequencies. This seems a more rational guess for $\tilde{\phi}(\omega)$ than ad hoc choices such as $\tilde{\phi}(\omega) = 0$ can provide. Note that (16) holds for any constant input polarization vector $\mathbf{e}_0$, i.e., independently of the distribution of power between the two arms, so that the output power will always be concentrated into the x-arm. This makes dropping the second polarization before detection a viable option. The price to pay for improved detection quality is a usually very small power penalty.

The simulation results for the above outlined partial measurement case (see, e.g., Fig. 11 in Section IV) revealed that the penalty due to the PMDC-induced chromatic dispersion is not neglible. Without taking into account alternative PMD penalty measures such as DOP or BER indicators at the receiver, a possible way to overcome this limitation is to obtain more complete phase information by switching the input polarization vector periodically between two non-collinear states. In principle, this would allow us to completely recover $U(\omega)$, avoid PMDC-induced chromatic dispersion to a large extent, and come close to the compensation results reproted for the full measurement case. These alternatives as well as a more detailed, experimentally supported investigation of the monitoring process are currently under investigation.

APPENDIX III
OPTIMIZATION: ALGORITHM DETAILS

We have tested several algorithmic approaches to the NLSQ minimization problem (7) where the PMDC is represented by the Jones matrix $U_{\mathrm{C}}(\omega)$ given in (5). For simplicity, we consider the full measurement case (8) and set $\mathbf{e}_0 = (1\ 0)^T$. As the preferred algorithm for solving (7), we have settled for the Levenberg-Marquardt (LM) algorithm implemented in Matlab's `lsqnonlin` for our tests. The main reason for this choice is its supposed efficiency and robustness in the pre-asymptotic range, i.e., in the first $10-20$ iterations, which we confirmed experimentally for our application by comparing it against Gauss-Newton-type methods in a large number of cases. Roughly speaking, in the LM algorithm, to find the $(k+1)$-th approximation x^{k+1} to the solution of (7) from the k-th approximation x^k, a new search direction d is produced by solving the linear system

$$[\mathbf{J}(x^k)^T\mathbf{J}(x^k) + \lambda I]d = -\mathbf{J}(x^k)^T\mathbf{F}(x^k), \tag{17}$$

followed by a line search. The search direction d depends on the parameter λ, and generally represents a compromise between the direction chosen in the Gauss-Newton method ($\lambda \to 0$) and a steepest descent direction ($\lambda \to \infty$). Thus, choosing λ appropriately in each iteration can enhance the decay of the functional value compared to both the Gauss-Newton and steepest decent methods.

The notation used in (17) is as follows. The real vector function $\mathbf{F}(x)$ (a column vector of length $4L$) is composed of the expressions for the real and imaginary parts of the two complex entries of the L vectors

$$r_i := U_{\mathrm{C}}(\omega_i)\tilde{\mathbf{e}}_i - (e^{j(\psi+\tau\omega_i)}\ 0)^T, \quad i = 1, \ldots, L,$$

and leads to the standard representation of the NLSQ functional $F(x) = \frac{1}{2}\|\mathbf{F}(x)\|^2$. The computation of $\mathbf{F}(x)$ for a particular set of parameters x essentially requires $\mathrm{O}(L+n)$ complex exponentials, and the equivalent of $\mathrm{O}(Ln)$ complex multiplications. $\mathbf{J}(x)$ is the $4L \times (2n+4)$ Jacobian matrix associated with $\mathbf{F}(x)$. A close look at the matrix product representing $U_{\mathrm{C}}(\omega)$ reveals that derivatives with respect to ϕ_k and θ_k can be obtained by replacing exactly one of the matrices in the product at each time. This allows us to compute the whole Jacobian $\mathbf{J}(x)$ by the equivalent of another $\mathrm{O}(Ln)$ complex multiplications, provided that intermediate results of the recursive computation for $\mathbf{F}(x)$ can be stored. In practice we observed that in our current implementation, a full $[\mathbf{F}(x), \mathbf{J}(x)]$ evaluation takes approximately the same time as three $\mathbf{F}(x)$ evaluations. We opted therefore for providing analytically computed Jacobians to `lsqnonlin`, which overall lead to a speedup. The line search along $x = x^k + \alpha d$, $\alpha \geq 0$, to find the next x^{k+1} requires additional $\mathbf{F}(x)$ evaluations. The determination of λ and solution of the linear system (17) seemingly account for only a small portion of the cpu-time per iteration step.

At input, a starting point x^0 needs to be provided which according to previous experience with choosing random starting points seems crucial for the success of `lsqnonlin` when applied to (7) for $n \geq 2$, as well as for avoiding sub-optimal solutions. Our current approach to the starting point problem is as follows:

a) First, we assume $\tau = \psi = 0$, and determine polynomials $p(z)$, $q(z)$ of degree n and the associated matrix $V_n(z)$ in (9) by solving two linear least-squares polynomial approximation problems such that

$$p(z_i) \approx z_i^{n/2}\tilde{u}_i^*, \quad q(z_i) \approx z_i^{n/2}\tilde{v}_i^*, \qquad z_i = e^{j\omega_i},$$

for all $i = 1, \ldots, L$. In practice, we have simply used the Matlab routine `polyfit`, and generally assumed that $L > n$. Recall that $\tilde{u}_i^*$, $\tilde{v}_i^*$ are the entries of $\tilde{\mathbf{e}}_i^*$, i.e., the first row in $U(\omega_i)^*$. Thus, $V_n(z) \approx U(\omega)^*$ in the control band. Since the control frequencies ω_i are not evenly spread over $[-\pi, \pi]$, a few additional control points (e.g., $z_0 = -1$) are added, and artificial data for

them are created by extrapolation of the available data $\tilde{e}_j$ in a periodic and unitary fashion. It turned out that this simple trick usually leads to polynomials $p(z), q(z)$ with a smaller deviation from the PC condition (10) outside the actual control region, and enhances the robustness of Step b) below.

b) Since the $V_n(z)$ from Step a) does not satisfy (10), we use the approximate factorization technique outlined in Appendix I. This gives a Jones matrix of degree n in the factorized form (14) which is hopefully still close to the $V_n(z)$ obtained in Step a), and thus close to $U(\omega)^*$ in the control band, too. By neglecting the matrix factor $Z_{\phi_{n+1}}$, we find $U_{\mathrm{C}}(\omega)$ as in (5), and accept its parameter set as the initial settings θ^0, ϕ^0.

c) In the last step, we compute the numbers $s_i = (1\ \ 0)U_C(\omega_i)\tilde{\mathbf{e}}_i$, $i = 1, \ldots, L$, and find values for ψ, τ such that $s_i \approx e^{j(\psi+\tau\omega_i)}$. This can be done by solving another small NLSQ problem, or (as was adopted for the simulation runs) by a linear fit for the unwrapped vector of phase angles corresponding to the s_i.

Finally, we note that (7) can be parametrized also in terms of the coefficients of the polynomials p_n, q_n which leads to a simple quadratic expession of the target functional in these new variables but requires adding quadratic constraints to ensure the PC condition. Although Matlab's `fmincon` [23] handles this formulation well, we did not pursue it further since the number of variables is roughly doubled which lead to less favorable running times.

PLACE PHOTO HERE

PLACE PHOTO HERE

PLACE PHOTO HERE

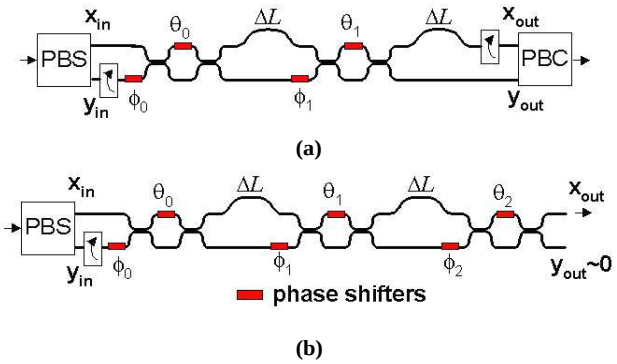


Fig. 1. PMDC architecture for $n = 2$: (a) with and (b) without PBC.

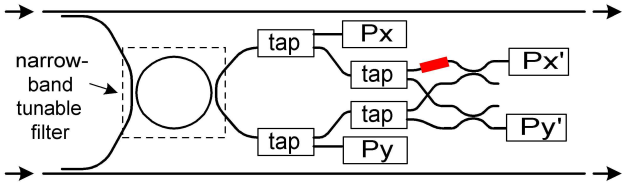


Fig. 2. Schematic view of the PMD monitor.

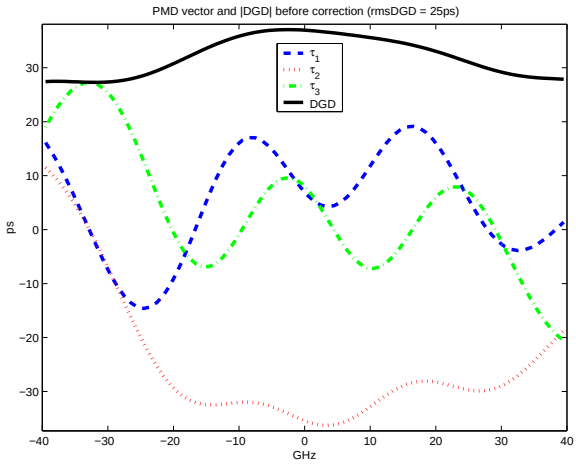


Fig. 3. Example with DGD of $\approx$ 30 ps: components of PMD vector and DGD before compensation

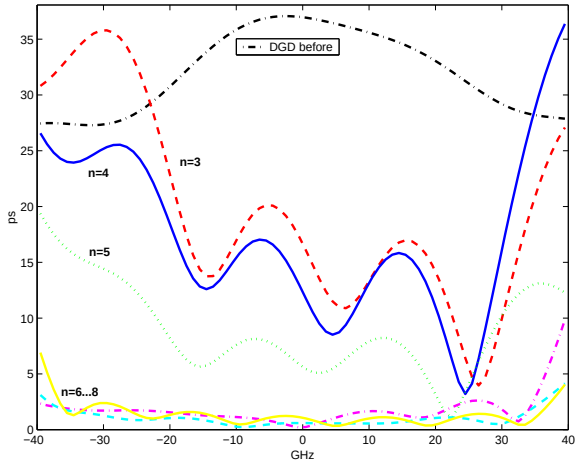


Fig. 4. Essential DGD after compensation for $n = 3, \ldots, 8$ by using (7)) for channel with $\tau_{\mathrm{DGD}}(\omega) \approx 30$ ps

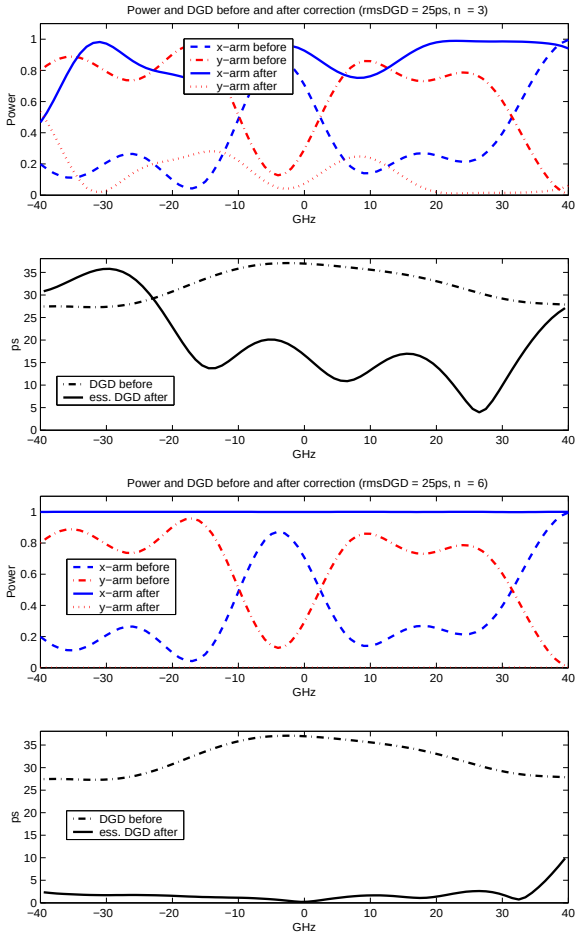


Fig. 5. Magnitude and DGD before and after correction with $n = 3$ and $n = 6$ stages for channel with $\tau_{\mathrm{DGD}}(\omega) \approx 30$ ps

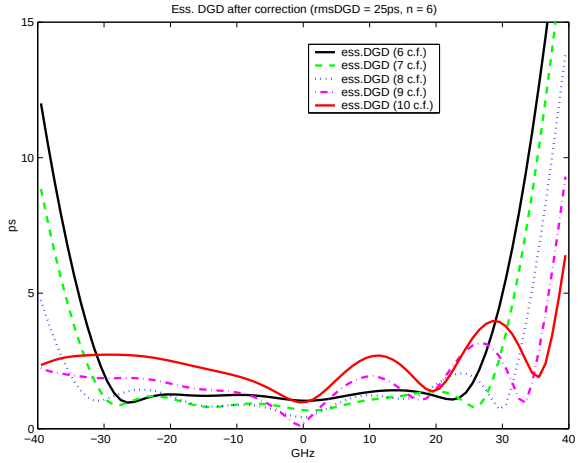


Fig. 6. Essential DGD after PMDC for $n = 6$ and Ω_C with $L = 6, \ldots, 10$ control freqencies spaced at 10 GHz

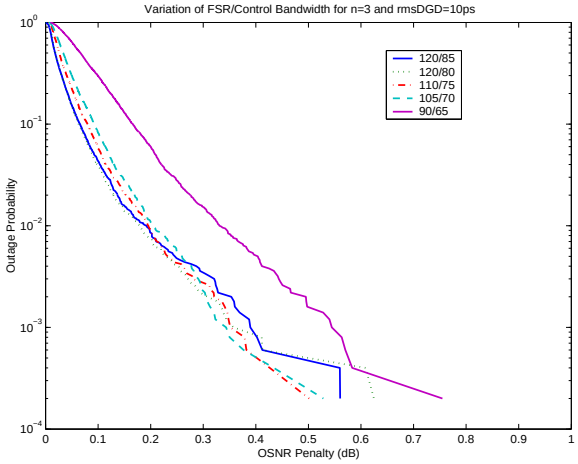


Fig. 7. Impact of varying FSR and control bandwidth for $n = 3$ and rmsDGD of 10 ps

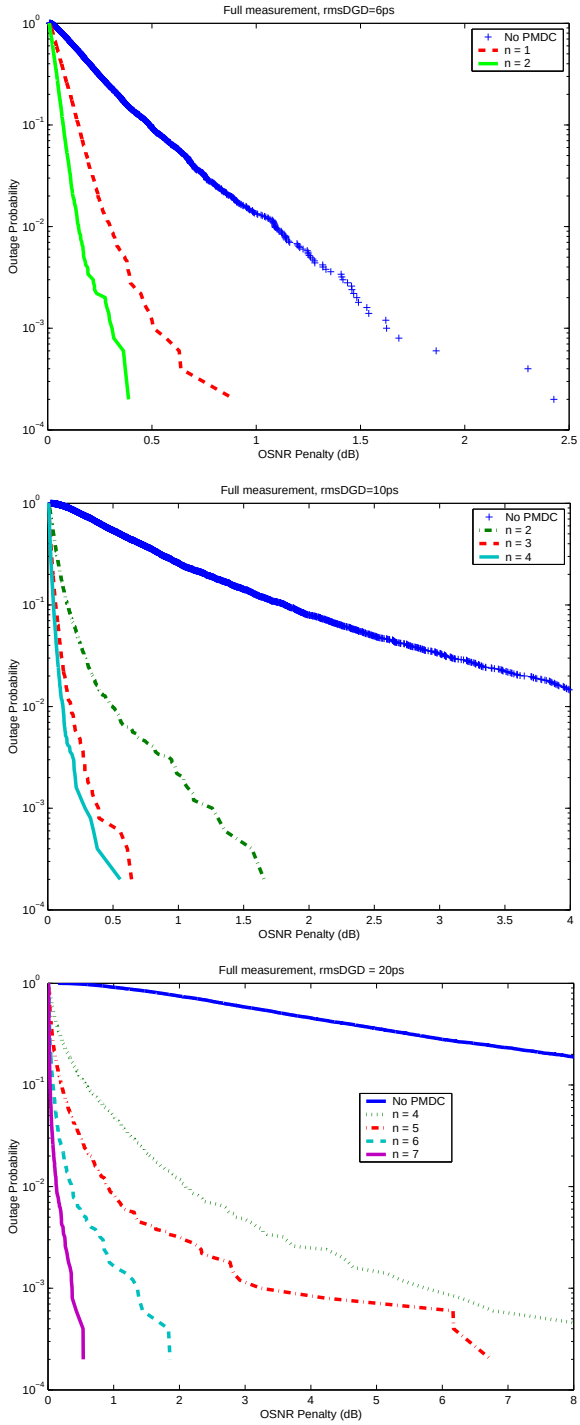


Fig. 8. Results for different number of fi lter stages and rms DGD values of 6 ps (top), 10 ps (center), and 20 ps (bottom)

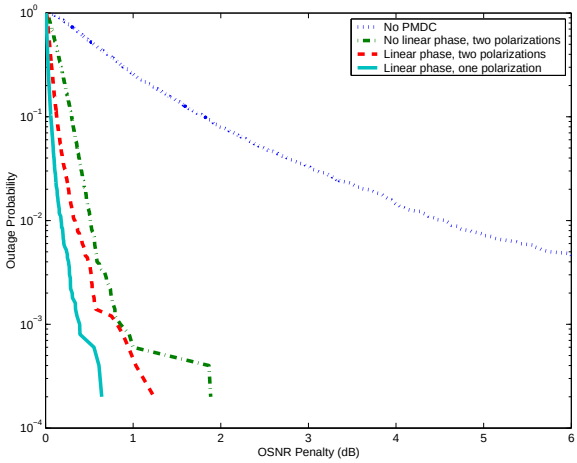


Fig. 9. Improvements from including τ and dropping the second polarization for rms DGD of 10 ps

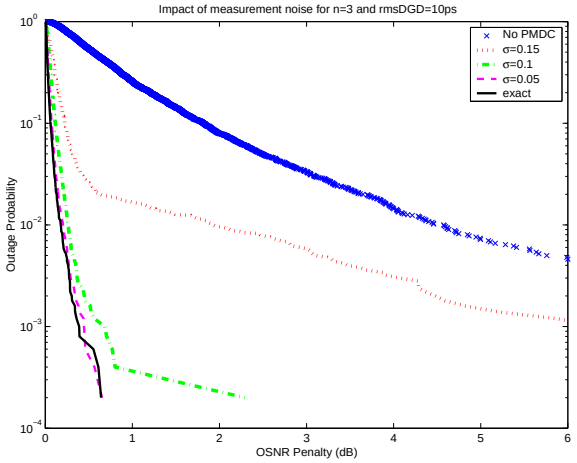


Fig. 10. Deterioriation of the PMDC result for different noise levels for rmsDGD of 10 ps and $n = 3$

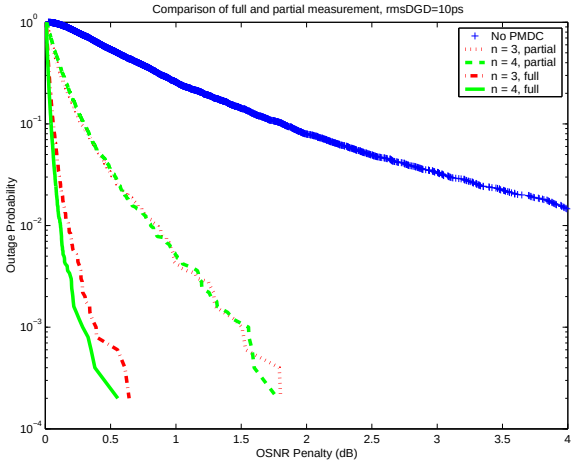


Fig. 11. Comparison of partial versus full measurements (rms DGD $\approx$ 10 ps, $n = 3$)