



# CHALMERS

## Chalmers Publication Library

### **Dispersion Compensation FIR Filter with Improved Robustness to Coefficient Quantization Errors**

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

**Journal of Lightwave Technology (ISSN: 0733-8724)**

Citation for the published paper:

Alireza, S. ; Fougstedt, C. ; Graell i Amat, A. et al. (2016) "Dispersion Compensation FIR Filter with Improved Robustness to Coefficient Quantization Errors". Journal of Lightwave Technology

Downloaded from: <http://publications.lib.chalmers.se/publication/241999>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

# Dispersion Compensation FIR Filter with Improved Robustness to Coefficient Quantization Errors

Alireza Sheikh, *Student Member, IEEE*, Christoffer Fougstedt, *Student Member, IEEE*,  
 Alexandre Graell i Amat, *Senior Member, IEEE*, Pontus Johannisson, *Member, OSA*,  
 Per Larsson-Edefors, *Senior Member, IEEE*, Magnus Karlsson, *Senior Member, IEEE, Fellow, OSA*

**Abstract**—In this paper, we propose a new finite impulse response (FIR) filter for chromatic dispersion compensation which is given in closed form. We identify a relation between the out-of-band gain and the in-band error when the filter is implemented with finite-precision arithmetic. In particular, a large out-of-band gain makes the filter more sensitive to coefficient quantization errors due to finite precision digital signal processing. To improve robustness to coefficient quantization errors, our proposed filter is designed based on confining the out-of-band gain. By means of simulations, we show that our filter outperforms other existing FIR filters. The performance gain improves with increasing modulation order and decreasing number of bits used to represent the filter taps.

## I. INTRODUCTION

CHROMATIC dispersion (CD), which causes pulse broadening and limits transmission data rate and reach, is one of the major impairments in coherent optical transmission systems [1], [2]. CD can be compensated in both the optical domain, e.g., by using a dispersion compensating fiber [3], and the digital domain, by means of digital signal processing. Finite impulse response (FIR) filters are particularly interesting, because they can be implemented efficiently in both the frequency and time domain. A first experimental demonstration in [4] showed that an FIR filter, designed by truncating the inverse Fourier transform of the ideal CD compensation filter, can compensate for accumulated CD over 89 km of a single-mode fiber by off-line processing. In [5], Savory proposed an FIR CD compensation filter, given in closed form, by performing direct sampling and truncation of the ideal CD compensation impulse response. The filter design in [5] did not consider the effect of pulse shaping but compensated CD over the entire bandwidth, including the band with no signal. Hence, it required a large number of filter taps [6]. Furthermore, such a filter has poor performance for higher-order modulations [6] and, as illustrated in Section IV, its

performance also degrades in the low-accumulated-dispersion regime, e.g., for transmission in short links.

CD compensation is one of the most power-hungry blocks in coherent systems [7], [8]. From an implementation point of view, it is preferable to design the filter with a small number of filter taps, since reducing the number of taps decreases the number of arithmetic operations, thus reducing the power consumption<sup>1</sup>. In [6], an FIR filter was proposed in the least-squares sense exploiting the fact that pulse shaping limits the effective bandwidth of the signal. The filter can then be designed to compensate the accumulated CD only in the signal band specified by the pulse shaping filter. It was shown in [6] that band-limited CD compensation allows to significantly reduce the nominal number of filter taps while still improving performance as compared to the filter in [5].

CD compensation is typically performed using a static CD compensation filter, which is usually accompanied by an adaptive filter to compensate for the residual CD, the polarization mode dispersion, and other unknown effects. CD can also be compensated for using only an adaptive filter. However, this is practical only for very short links which require a small number of taps (e.g., 20 km with 9 taps [9]). Here, we consider longer links. For (longer) links with relatively large accumulated CD, the number of taps required for the adaptive equalizer is large (e.g., 600 km requires 242 taps [9]), which leads to slow convergence speed in updating the taps and much higher complexity than the static filter [9]. Furthermore, in this scenario an adaptive filter consumes high power [10]. Therefore, our focus is on static CD compensation.

Previous works on CD compensation implicitly assume that filter taps are implemented using floating-point precision. In practice, however, filter taps must be implemented using finite-precision arithmetic<sup>2</sup>. Coefficient quantization alters the values of the coefficients, thus introducing quantization errors which, in turn, change the frequency response of the filter. As a result, a good filter design for floating-point precision may not be adequate if the filter is implemented using finite precision.

In this paper, we consider the design of FIR filters by taking into consideration the fact that filter taps are implemented with

This work was financially supported by the Knut and Alice Wallenberg Foundation.

A. Sheikh and A. Graell i Amat are with the Department of Signals and Systems, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (e-mail: {asheikh,alexandre.graell}@chalmers.se).

C. Fougstedt and P. Larsson-Edefors are with the Department of Computer Science and Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (e-mail: {chrfou,perla}@chalmers.se).

P. Johannisson and M. Karlsson are with the Photonics Laboratory, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (e-mail: {pontus.johannisson,magnus.karlsson}@chalmers.se).

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

<sup>1</sup>Note that a shorter filter decreases also the induced latency. However, in current coherent optical systems the latency induced by the FIR CD compensation filter is typically less than the latency due to transmission over the fiber and soft-decision forward error correction. Therefore, it does not have a dominant effect in the overall system latency.

<sup>2</sup>We remark that floating-point precision is also finite, e.g., MATLAB uses 64 bits for double-precision floating point. In this paper “finite precision” means a small number of bits (3–5 bits) used for filter coefficient quantization.

finite-precision arithmetic. The paper is an extension of our previous work [11]. We show that there is a relation between the FIR filter out-of-band gain and coefficient quantization errors, in the sense that a larger out-of-band gain leads to higher coefficient quantization errors. To reduce the impact of coefficient quantization errors, we propose an FIR filter design that confines the out-of-band gain. The filter is designed based on solving a constrained least-squares optimization problem, and the filter taps are derived in closed form. By means of simulations, we will show that our proposed filter achieves better performance in terms of bit error rate (BER) as compared to other FIR filters when finite-precision is used. The performance gain becomes more substantial as the modulation order increases. We also analyze the effect of the filter length in the performance and show that reducing the number of filter taps leads to a higher out-of-band gain, thus to higher coefficient quantization errors. Therefore, a significant reduction in the number of filter taps as advocated in [6] yields a severe penalty in terms of BER when finite precision is used. Nevertheless, we show that a reduction factor up to 40 percent can be achieved with little penalty on the performance.

Our proposed FIR filter is suitable for both time-domain and frequency-domain implementations. For short links, time-domain CD compensation can be more power efficient [12]. For instance, the time-domain implementation of our proposed filter has lower power consumption than the frequency-domain equalization (FDE) for fiber lengths up to 150 km [12]. On the other hand, for long links, where the accumulated CD is large, it is more power efficient to use FDE methods [13]. We show that for long links the proposed filter can be used as the component of the FDE configuration using overlap-save method [14] with better performance than previously proposed filters. In this paper we do not discuss implementation issues of the filter. However, the interested reader is referred to [12].

## II. PRELIMINARIES

CD is modeled as an all-pass filter with frequency response [6, Eq. (4)]

$$H_{\text{CD}}(e^{j\omega T}) = e^{-jM(\omega T)^2}, \quad M = \frac{D\lambda^2 L}{4\pi c T^2}, \quad (1)$$

where  $D$  is the dispersion parameter,  $T$  is the sampling interval,  $L$  is the fiber length,  $\lambda$  is the operating wavelength,  $c$  denotes the speed of light, and  $\omega T \in [-\pi, \pi]$  is the digital frequency. The CD can be ideally compensated by an all-pass static filter with frequency response

$$H_{\text{CD}}^{-1}(e^{j\omega T}) = e^{jM(\omega T)^2}. \quad (2)$$

By taking the inverse Fourier transform of (2), the ideal CD compensation impulse response can be derived as [5, Eq. (6)]. In [5], by direct sampling and truncation of the continuous CD compensation filter impulse response, the  $n$ th filter tap is derived as

$$h_n^{\text{DS}} = \sqrt{\frac{j}{4M\pi}} e^{-j\frac{n^2}{4M}} \quad (3a)$$

$$-\frac{N_{\text{max}} - 1}{2} \leq n \leq \frac{N_{\text{max}} - 1}{2}, \quad (3b)$$

Table I  
SIMULATION PARAMETERS

$L$ (km)	$B$ (Gbaud)	$N$ ( $0.6N_{\text{max}}$ )	$D$ (ps/nm/km)	$M$	$\lambda$ (nm)	$\rho$	$G$
250	28	61	16	8.00	1550	0.25	2

where  $N_{\text{max}}$  is the maximum number of filter taps so that aliasing does not occur, in the sense that a filter with more taps would suffer from aliasing. Aliasing occurs when the angular frequency of the ideal CD compensation impulse response exceeds the Nyquist frequency determined by the sampling rate.  $N_{\text{max}}$  is given by [5, Eq. (9)]

$$N_{\text{max}} = 2 \lfloor 2\pi |M| \rfloor + 1, \quad (4)$$

where  $\lfloor x \rfloor$  gives the greatest integer smaller than or equal to  $x$  and  $|\cdot|$  indicates absolute value. In the following we will refer to this filter in [5] (eqs. (3a) and (3b)) as the direct sampling (DS) filter.

Pulse shaping at the transmitter is used to reduce the inter-symbol interference and makes the signal band limited. Square-root raised-cosine (SRRC) pulses are popular for pulse shaping and matched filtering due to their anti-aliasing characteristics [15]. The frequency response of an SRRC pulse is  $\sqrt{f(\omega T)}$ , where  $f(\omega T)$  is given in (5) at the top of the next page, where  $G$  denotes the number of samples per symbol and  $\rho$  is the roll-off factor [15].

In [6], a least-squares optimization problem was proposed to compensate for the CD in the signal band. The considered objective function was the energy of the complex error between the frequency response of the ideal CD compensation filter and the proposed FIR filter in the signal band. We refer to this filter as the least-squares band-limited (LS-BL) filter. The filter taps can be found in [6, Eqs. (9)–(13)]. The LS-BL filter requires an adjustment parameter,  $\varepsilon$ , to prevent ill-conditioned matrices in the procedure of filter tap calculations. If the operations are done in double-precision floating point, the objective function is monotonically decreasing by decreasing the adjustment parameter until  $\varepsilon = 10^{-15}$ , where numerical problems arise. To avoid numerical problems,  $\varepsilon = 10^{-14}$  was suggested in [6].

### A. Motivation for a New Filter Design

Fig. 1(a) depicts the amplitude response of the LS-BL filter with SRRC pulse shaping and parameters given in Table I for  $N_{\text{max}}$  taps (blue curve) and  $0.6N_{\text{max}}$  taps (red curve). The filter taps are implemented using floating-point precision. We refer to the in-band frequencies as  $[-0.625\pi, 0.625\pi]$  (see (5)). It can be observed that a decrease in the number of filter taps leads to an increase of the out-of-band gain. In particular, for  $0.6N_{\text{max}}$  taps, it reaches 70 dB. However, for both  $N_{\text{max}}$  and  $0.6N_{\text{max}}$  taps the LS-BL filter can compensate for the accumulated CD [6].

In Fig. 1(b), we plot the amplitude response of the same filter using 4 bits to implement the filter taps. The procedure of quantizing the filter taps is explained in detail in Section III-B. As can be observed, for  $N_{\text{max}}$  taps, finite-precision arithmetic adds some wiggling on the in-band response of the filter, but

$$f(\omega T) = \begin{cases} 1 & |\omega T| \leq \frac{1-\rho}{G}\pi \\ \frac{1}{2} \left[ 1 + \cos \left[ \frac{1}{2\rho} (G|\omega T| - (1-\rho)\pi) \right] \right] & \frac{1-\rho}{G}\pi \leq |\omega T| \leq \frac{1+\rho}{G}\pi \end{cases} \quad (5)$$

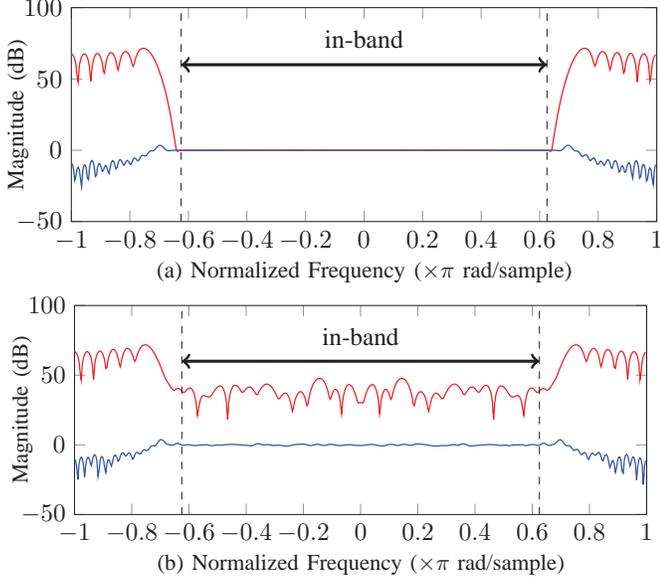


Figure 1. Amplitude response of LS-BL filter with  $N_{\max}$  (blue curve) and  $0.6N_{\max}$  taps (red curve) a) in double precision floating point b) in fixed point using 4-bit word length.

does not destroy the filter response. On the other hand, if  $0.6N_{\max}$  taps are used the in-band response of the filter is significantly distorted. In this case, the filter cannot compensate for the accumulated CD. A high quantization error translates into a lot of distortion in the frequency domain.

Let  $\xi_q$  be the overall quantization error between the filter implemented in floating-point precision,  $\mathbf{h}$ , and the filter implemented using finite-precision arithmetic,  $\mathbf{h}_q$ , normalized by the number of filter taps,  $N$ ,

$$\xi_q = \frac{\|\mathbf{h} - \mathbf{h}_q\|^2}{N}, \quad (6)$$

where  $\|\mathbf{h}\|^2 \triangleq \mathbf{h}^H \mathbf{h}$ , and  $[\cdot]^H$  denotes hermitian conjugation. Then,  $\xi_q = 197$  for the filters with  $0.6N_{\max}$  filter taps (red curves) and  $\xi_q = 0.00004$  for the filters with  $N_{\max}$  filter taps (blue curves). In Section III-B we show that for finite-precision, a higher out-of-band gain translates into a higher quantization error. Therefore, in the next section, we propose a filter design, in a constrained least-squares sense, that confines the out-of-band gain.

### III. PROPOSED FIR CD COMPENSATION FILTER DESIGN

We denote the  $n$ th filter tap by  $h_n$ . The discrete Fourier transform (DFT) of the filter can be written as

$$H(e^{j\omega T}) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} h_n e^{-jn\omega T}, \quad (7)$$

where  $N$  is assumed to be an odd number. The aim is designing a filter to have a frequency response as similar as

possible to the ideal CD compensation filter within the signal band. We define the normalized average filter response error within the signal band as

$$\begin{aligned} \xi_s &= \frac{\int_{\Omega_s} |H(e^{j\omega T}) - H_{\text{CD}}^{-1}(e^{j\omega T})|^2 d(\omega T)}{\int_{\Omega_s} d(\omega T)} \\ &= \frac{1}{\Omega_2 - \Omega_1} \int_{\Omega_1}^{\Omega_2} |H(e^{j\omega T}) - H_{\text{CD}}^{-1}(e^{j\omega T})|^2 d(\omega T), \end{aligned} \quad (8)$$

where  $\Omega_1$  and  $\Omega_2$  denote the boundary frequencies of the signal band  $\Omega_s$ . Assuming SRRC pulse shaping,  $\Omega_2 = -\Omega_1 = (1+\rho)\pi/G$ . The normalized average out-of-band gain of the filter is

$$\begin{aligned} \xi_o &= \frac{\int_{\Omega_o} |H(e^{j\omega T})|^2 d(\omega T)}{\int_{\Omega_o} d(\omega T)} = \frac{1}{2\pi + \Omega_1 - \Omega_2} \\ &\quad \left[ \int_{-\pi}^{\Omega_1} |H(e^{j\omega T})|^2 d(\omega T) + \int_{\Omega_2}^{\pi} |H(e^{j\omega T})|^2 d(\omega T) \right], \end{aligned} \quad (9)$$

where  $\Omega_o$  denotes the spectral bands with no signal content. We propose an optimization problem to minimize  $\xi_s$  while confining the out-of-band gain  $\xi_o$ . Considering  $\mathbf{h} = [h_{-(N-1)/2}, \dots, h_0, \dots, h_{(N-1)/2}]^T$ , where  $[\cdot]^T$  denotes the transpose, the optimization problem can be formulated as

$$\tilde{\mathbf{h}} = \arg \min_{\mathbf{h}} \xi_s \quad (10)$$

$$\text{subject to } \xi_o \leq \xi_{o,\max}, \quad (11)$$

where  $\xi_{o,\max}$  is a selected threshold on the out-of-band gain. The optimization variables are the filter tap coefficients and  $\xi_{o,\max}$ . We are interested in solving (10) with the constraint in (11) when quantization errors are introduced due to the implementation of the filter taps using finite-precision arithmetic.<sup>3</sup> To reduce the sensitivity to coefficient quantization errors, we present a heuristic two-step algorithm as follows. We will refer to this filter as the least-squares constraint optimization (LS-CO) filter.

- Step A (Filter taps calculation). Exclude  $\xi_{o,\max}$  from the optimization variables. Then, calculate the filter taps as a function of  $\xi_{o,\max}$ .
- Step B (Threshold selection). Find the threshold  $\xi_{o,\max}$  that minimizes the in-band error in the presence of coefficient quantization errors.

In the following, we describe each step in detail.

#### A. Filter tap calculations

By considering  $\xi_{o,\max}$  as a parameter and not as an optimization variable, the optimization problem is convex and can

<sup>3</sup>Note that (10) without the constraint in (11) corresponds to the optimization problem for the LS-BL filter.

be classified as a constrained least-squares problem, which can be solved using the Lagrangian method [16].

*Theorem 1:* For a given  $\xi_{o,\max}$ , the optimal filter tap coefficients  $\tilde{\mathbf{h}} = [\tilde{h}_{-(N-1)/2}, \dots, \tilde{h}_0, \dots, \tilde{h}_{(N-1)/2}]^T$  obtained by solving the optimization problem (10) are

$$\tilde{\mathbf{h}} = \mathbf{Q}^{-1}\mathbf{v}, \quad (12)$$

where  $\mathbf{Q}$  is an  $N \times N$  matrix and  $\mathbf{v}$  is an  $N \times 1$  vector  $\mathbf{v} = [v_{-(N-1)/2}, \dots, v_0, \dots, v_{(N-1)/2}]^T$  with elements given in (13) and (14), respectively, at the top of the next page. The Lagrangian parameter  $\lambda$  in (13) can be calculated by solving

$$\frac{2\pi\tilde{\mathbf{h}}^H\tilde{\mathbf{h}} + \Omega_1 - \Omega_2}{2\pi + \Omega_1 - \Omega_2} = \xi_{o,\max}. \quad (15)$$

*Proof:* The proof is given in Appendix A. ■

The function  $\text{erfz}$  in (14) is the error function of complex arguments defined as [17, Eq. (7.1.5)]

$$\text{erfz}(x) = \frac{2x}{\sqrt{\pi}} \left( 1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{n!(2n+1)} \right). \quad (16)$$

An efficient numerical method for computing (16) can be found in [17].

### B. Threshold selection

The in-band error defined in (8) only includes the spectral bands defined by the pulse shaping. To optimize  $\xi_{o,\max}$ , we include the effect of pulse shaping. Let  $G_{\text{TX}}(\omega T)$  be the frequency response of the transmit pulse, and assume a matched filter at the receiver. Then,  $\xi_{o,\max}$  is selected such that

$$\bar{\xi}_s = \frac{1}{\Omega_2 - \Omega_1} \int_{\Omega_1}^{\Omega_2} |H(e^{j\omega T}) - H_{\text{CD}}^{-1}(e^{j\omega T})|^2 |G_{\text{TX}}(\omega T)|^2 d(\omega T), \quad (17)$$

is minimized (cf. (8)). In Fig. 2, we plot  $\bar{\xi}_s$  as a function of  $\xi_{o,\max}$  for SRRC pulse shaping. The filter taps are derived according to (12)–(15) and floating-point precision is assumed. As can be seen,  $\bar{\xi}_s$  decreases with  $\xi_{o,\max}$ . Therefore, for floating-point precision, a larger  $\xi_{o,\max}$  leads to a better filter.

Consider now a more realistic system where the filter taps are implemented using finite-precision arithmetic. In particular, we are interested in the filter tap coefficients whose quantized version minimize the in-band error which, in turn, leads to a better filter in terms of CD compensation. The filter taps are quantized in the following way. The filter tap vector is scaled down by a normalization factor, which is defined as the maximum value of the real and imaginary components of the filter tap coefficients. Then, the scaled filter taps are quantized linearly according to the number of bits used for quantization. Finally, the vector is scaled up.

In Fig. 3, we show  $\tilde{\xi}_q = \frac{\|\tilde{\mathbf{h}} - \tilde{\mathbf{h}}_q\|^2}{N}$  as a function of  $\xi_{o,\max}$  for the parameters in Table I and different number of bits used for quantization. Regardless of the word length, a large out-of-band threshold makes the filter more sensitive to coefficient quantization errors, hence  $\tilde{\xi}_q$  increases. This can be seen by taking a look at the Lagrangian function of (10) in (18). A larger out-of-band threshold leads to a smaller Lagrangian parameter. On the other hand, with respect to (12) and (13),

Table II  
LIST OF ACRONYMS

<b>DS:</b>	direct sampling	<b>LS-BL:</b>	least-squares band-limited
<b>LS-FB:</b>	least-squares full band	<b>LS-CO:</b>	least-squares constrained optimization

which depend on the Lagrangian parameter, one can expect that, on average, the values of the filter tap coefficients increase with decreasing Lagrangian parameter. Therefore, for a given word length, the induced quantization errors will also increase on average. Based on Fig. 2 and Fig. 3, we can infer that the threshold for the out-of-band gain is a tuning parameter which establishes a tradeoff between coefficient quantization errors (out-of-band gain) and in-band error. Assuming a certain word length for filter taps, if we sweep the threshold and then depict the corresponding modified in-band error, we can select the best threshold to minimize the modified in-band error. In Fig. 4,  $\bar{\xi}_s$  is depicted as a function of  $\xi_{o,\max}$  for the same parameters and different word lengths. As can be seen, the optimum threshold varies for different word length. The best threshold is  $-0.54$  dB,  $-0.13$  dB, and  $-40$  dB for 5, 4, and 3 bit word length, respectively. We remark that if high precision is possible, the coefficient quantization errors can be neglected. In this case, the optimal solution to (10) is achieved for  $\xi_{o,\max} \rightarrow \infty$ , i.e., the constraint on the out-of-band gain in (11) is lifted. As a consequence, the optimization problem is identical to that of the LS-BL filter. In other words, for floating-point precision the LS-CO filter boils down to the LS-BL filter.

## IV. SIMULATION RESULTS

The simulation chain is shown in Fig. 5, where  $g_{\text{TX}}(t)$  is the transmit pulse (pulse shaping),  $g_{\text{RX}}(t)$  the matched filter, and  $z(n)$  is additive white Gaussian noise. In our simulations, we consider a single channel, SRRC pulse shaping, and rectangular constellations with Gray mapping. The fiber parameters are summarized in Table I. We consider a short link scenario, where FIR filters are of practical interest (c.f. Section I). For short links, the system is in the linear regime and is limited by amplified spontaneous emission noise. Therefore, the nonlinearity can be reasonably neglected. (In long-haul links, the nonlinearity is not negligible and CD can be compensated in the frequency domain.) Since the goal of this work is to investigate the effect of finite precision on the performance of the FIR CD compensation filter, we neglect the other non-idealities. In particular, we do not include laser phase noise and assume floating-point precision of the analog-to-digital converter, i.e., no additional quantization errors to the input signal of the CD compensation filter are added. Furthermore, the number of samples in the digital pulse shaping is large enough so that no errors arise from pulse shaping and matched filtering. The acronyms stand for different filters is summarized in Table II.

As discussed in Section II, the LS-BL filter has an adjustment parameter,  $\varepsilon$ , which was optimized in [6] assuming floating-point arithmetic. For the sake of fairness in our comparisons, we performed an exhaustive search to find the best  $\varepsilon$  for a given finite precision. Our simulations show

$$Q_{m,n} = \begin{cases} \frac{2\pi(\lambda+1)+(\lambda+1)\Omega_1-(\lambda+1)\Omega_2}{2\pi+\Omega_1-\Omega_2} & \text{if } m = n \\ \frac{\lambda}{j(m-n)(2\pi+\Omega_1-\Omega_2)} [e^{j(m-n)\Omega_1} - e^{j(m-n)\Omega_2}] + \frac{1}{j(m-n)(\Omega_1-\Omega_2)} [e^{j(m-n)\Omega_1} - e^{j(m-n)\Omega_2}] & \text{if } m \neq n \end{cases} \quad (13)$$

$$v_m = \frac{e^{-j\left(\frac{m^2}{4M} + \frac{3\pi}{4}\right)}}{2(\Omega_2 - \Omega_1)} \sqrt{\frac{\pi}{M}} \left\{ \operatorname{erfz} \left[ \frac{e^{-j\frac{\pi}{4}}(2M\Omega_1 + m)}{2\sqrt{M}} \right] - \operatorname{erfz} \left[ \frac{e^{-j\frac{\pi}{4}}(2M\Omega_2 + m)}{2\sqrt{M}} \right] \right\}, \quad (14)$$

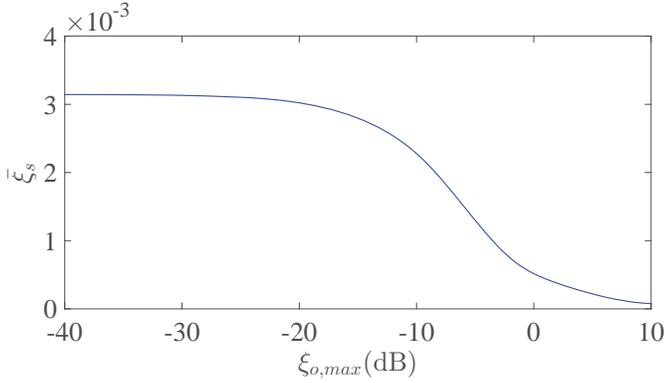


Figure 2.  $\bar{\xi}_s$  versus threshold for a system with parameters listed in Table I and filter coefficients with floating-point precision.

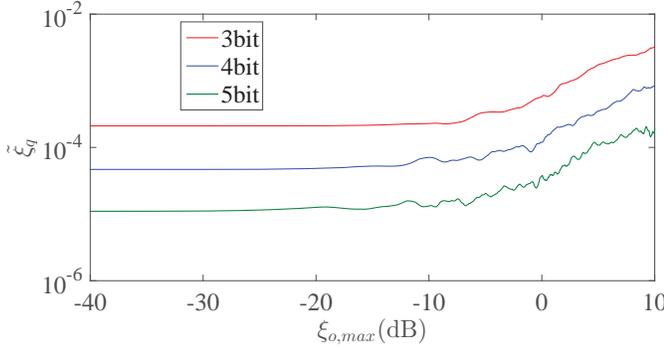


Figure 3.  $\bar{\xi}_q$  versus threshold for a system with parameters listed in Table I with 3-, 4-, and 5-bit filter coefficient word length.

that, for finite precision, the LS-BL filter is highly sensitive to the adjustment parameter and a small deviation from the optimized value significantly degrades the performance. Also it takes fairly long simulation time to find the best adjustment parameter for a given system parameters. On the other hand, if CD is compensated in the full band, the adjustment parameter is not required [6]. We refer to the filter in [6] that compensates CD in the full band as the least-squares full band (LS-FB) filter.

In Fig. 6, we compare the BER performance of our proposed filter (referred to as the least-squares constrained optimization (LS-CO) filter in this and the following figures) with the LS-BL and LS-FB filters in [6] and the DS filter in [5] for QPSK, and 8-QAM modulations. Fig. 7 gives BER results for 16-QAM and 64-QAM. The filter taps are quantized using 4 bits. The proposed filter yields better performance compared to the other filters, i.e., it is more robust to coefficient quantization errors. Furthermore, the performance gain increases for higher modulation order. As an example, for  $\text{BER} = 10^{-3}$ , the performance gain of the LS-CO filter over the LS-BL filter is 0.07 dB, 0.21 dB, and 0.44 dB for QPSK, 8-QAM, and 16-QAM,

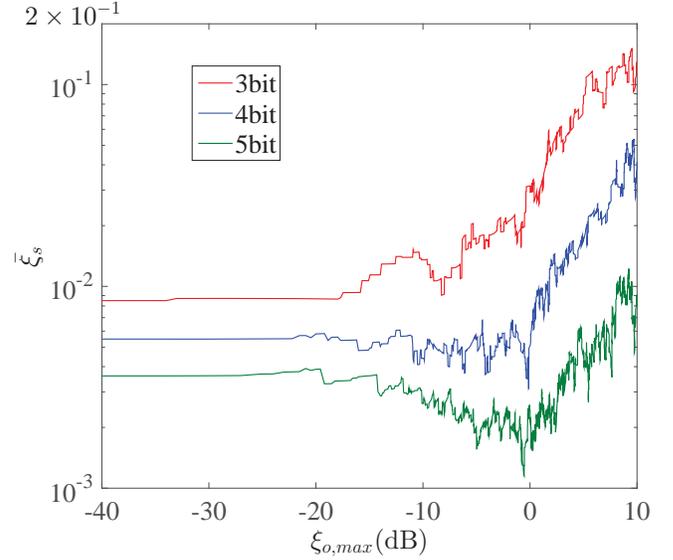


Figure 4.  $\bar{\xi}_s$  versus threshold for a system with parameters listed in Table I with 3-, 4-, and 5-bit filter coefficient word length.

respectively. The higher gain for higher order modulations is expected: Since constellation points are closer for higher-order modulations, the received signal is more sensitive to residual uncompensated CD after FIR CD compensation. The LS-CO filter has lower residual CD as compared to the other filters, thus it yields a higher performance improvement for higher modulation orders.

To evaluate the effect of the finite precision on the BER performance, in Fig. 8 we plot BER results for 16-QAM using 3-, 4-, and 5-bit precision. The DS and LS-FB filters perform very poorly for all considered finite precisions. Also, the LS-BL and our proposed filter show very poor performance for 3-bit precision. A significant improvement is observed increasing to 4-bit precision, and performance close to that of the floating-point implementation is observed using 5-bit precision. In all cases, our proposed filter gives the best performance. However, we point out that the performance gain decreases with increasing the number of bits used for precision, since the coefficient quantization error decreases. For a  $\text{BER} = 10^{-3}$ , the penalty due to the fixed-point implementation compared to floating-point implementation for 4- and 5-bit precision is 0.74 dB and 0.19 dB for our proposed filter and 1.17 dB and 0.34 dB for LS-BL filter, respectively. Ultimately (for floating-point precision) our proposed filter and the LS-BL filter become identical (see Section III).

In Figs. 9, 10 and 11 we investigate the effect of reducing the number of filter taps on the performance. In Fig. 9 we plot the required SNR at the input of the CD compensation filter,  $\text{SNR}_{\text{in}}$ , to achieve  $\text{BER} = 10^{-3}$  using 4 bits for precision as a

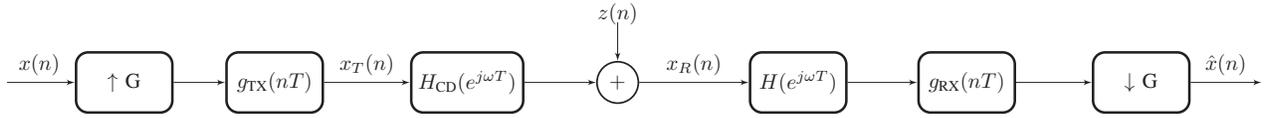


Figure 5. Simulation chain consists of up-sampling to  $G$  samples per symbol, pulse shaping  $g_{TX}(nT)$ , CD channel, FIR CD compensation filter, matched filter  $g_{RX}(nT)$ , and down-sampling.

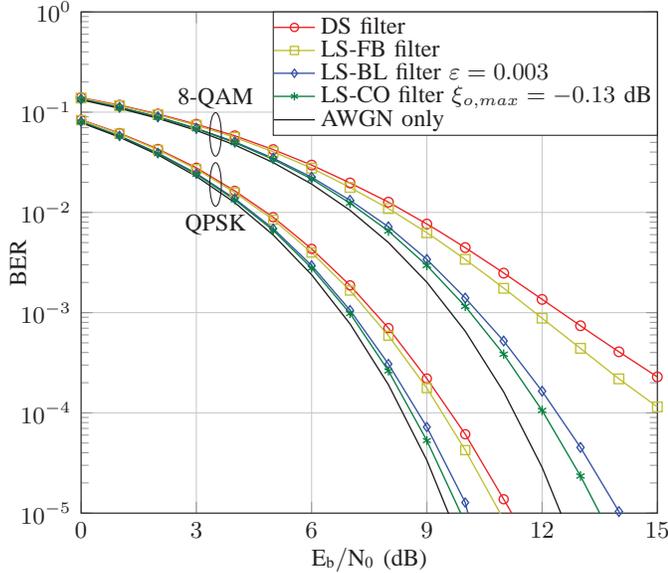


Figure 6. BER performance of different filters for QPSK, and 8-QAM modulations with 4-bit filter coefficient word length.

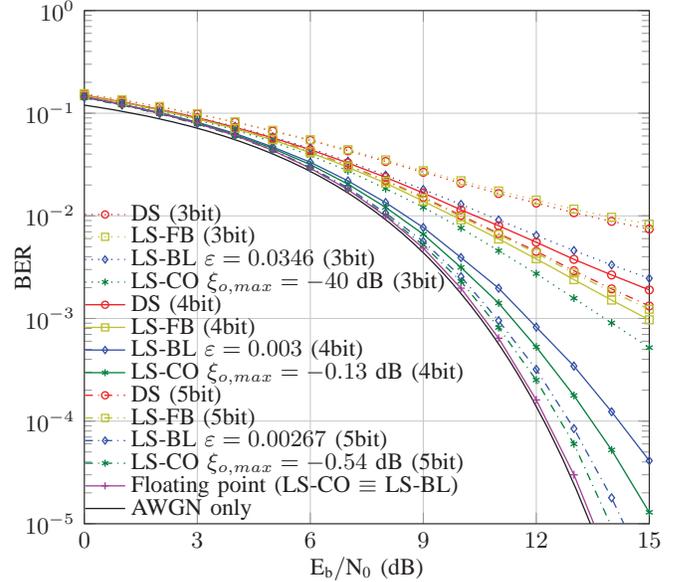


Figure 8. BER performance of different filters for 16-QAM modulations with 3-, 4-, and 5-bit filter coefficient word length.

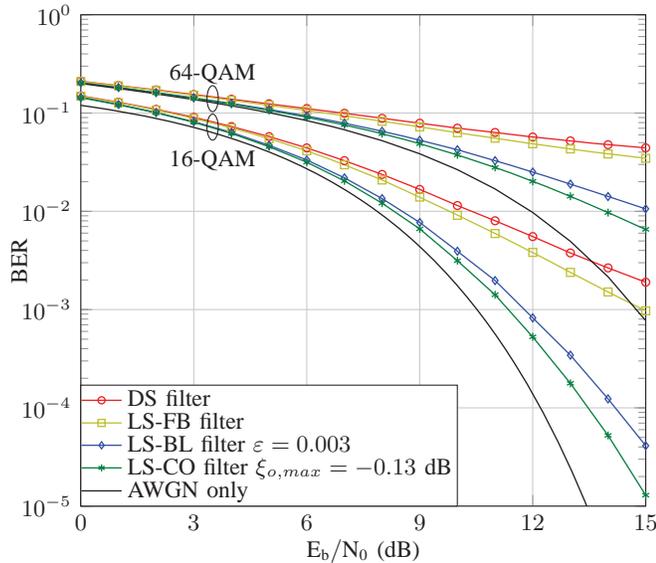


Figure 7. BER performance of different filters for 16-QAM, and 64-QAM modulations with 4-bit filter coefficient word length.

function of the symbol rate (from the low-dispersion regime, 7 Gbaud, to the high-dispersion regime, 56 Gbaud). The modulation format is 16-QAM and the remaining parameters are taken from Table I, except the symbol rate, which is varying here. The number of filter taps is set to  $0.6N_{\max}$ , corresponding to 5, 15, 61, 137, and 241 taps for 7 Gbaud, 14 Gbaud, 28 Gbaud, 42 Gbaud, and 56 Gbaud, respectively. It can be seen that our proposed filter requires lower  $\text{SNR}_{\text{in}}$  for all symbol rates as compared to the other filters. Note that by increasing the symbol rate,  $\text{SNR}_{\text{in}}$  decreases since the

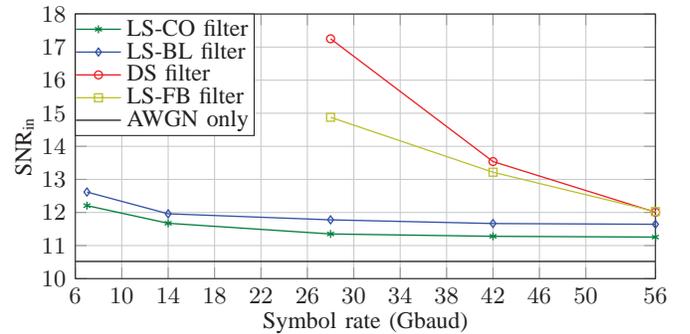


Figure 9. CD compensation required input SNR ( $\text{SNR}_{\text{in}}$ ) versus symbol rate using 16-QAM modulation and 4-bit filter coefficient word length. The number of filter taps is set to  $0.6N_{\max}$ , corresponding to 5, 15, 61, 137, and 241 taps for 7 Gbaud, 14 Gbaud, 28 Gbaud, 42 Gbaud, and 56 Gbaud, respectively.

number of filter taps increases, which leads to an increase in the number of variables in solving (10) and better performance should be expected. The same behavior is observed for all other filters. Note also that the performance improvement saturates in the high-dispersion regime, since in this case the number of filter taps is enough to compensate the CD and the performance is limited by coefficient quantization errors. The DS and LS-FB filters show poor performance in the low-dispersion regime. More precisely, for symbol rates from 7 to 14 Gbaud, both have an error floor higher than  $10^{-3}$  which makes the  $\text{SNR}_{\text{in}}$  infeasible.

In Figs. 10 and 11 we plot the required  $\text{SNR}_{\text{in}}$  to achieve  $\text{BER} = 10^{-3}$  as a function of the number of filter taps (normal-

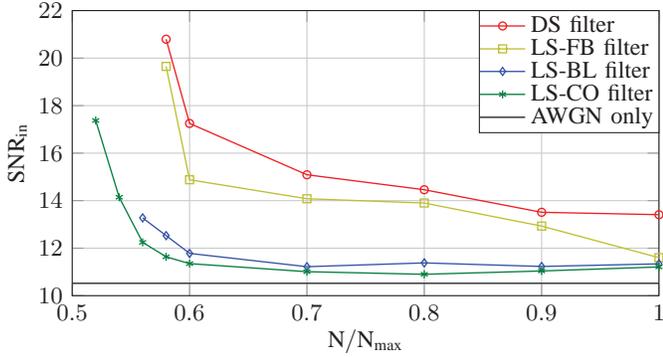


Figure 10. CD compensation required input SNR ( $\text{SNR}_{\text{in}}$ ) versus tap numbers using 16-QAM modulation with 4-bit filter coefficient word length.

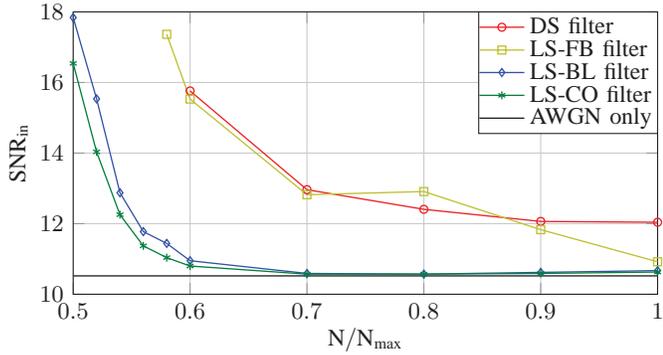


Figure 11. CD compensation required input SNR ( $\text{SNR}_{\text{in}}$ ) versus tap numbers using 16-QAM modulation with 5-bit filter coefficient word length.

ized by  $N_{\text{max}}$  for 4-bit and 5-bit precision, respectively. Again, 16-QAM is considered and the system parameters except the number of taps, which is varying here, are taken from Table I. For a given number of filter taps, our proposed filter requires lower  $\text{SNR}_{\text{in}}$  as compared to the other filters. As a reference, we also plot the required  $\text{SNR}_{\text{in}}$  for a system with AWGN only. As expected, the gap compared to the AWGN performance reduces for increasing precision and number of filter taps. For all values of  $N/N_{\text{max}}$ , our proposed filter and the LS-BL filter require significantly lower  $\text{SNR}_{\text{in}}$  as compared to the DS and LS-FB filters, with the proposed filter achieving the best performance. We remark that increasing the number of filter taps above  $N_{\text{max}}$  degrades the performance of the DS filter due to aliasing (see Section II) but can improve the performance of the other filters, especially in the low accumulated CD regime, at the expense of a higher latency and power consumption. We also remark that the interesting region in Figs. 10 and 11 is the left portion of the horizontal axis, since a lower number of taps reduces power consumption of the filter implementation. We also observe that for a number of taps smaller than  $0.6N_{\text{max}}$  the performance degrades significantly. This is in contrast to the observation (for floating-point precision) in [6], where it was claimed that a reduction factor of up to 52 percent over the maximum number of filter taps incurs little performance degradation. According to the results in Figs. 10 and 11, however, when finite precision is considered, the reduction factor should be pushed backward to 40 percent.

In Fig. 12 we compare the performance of the FDE configuration using the LS-CO and DS filters as the component

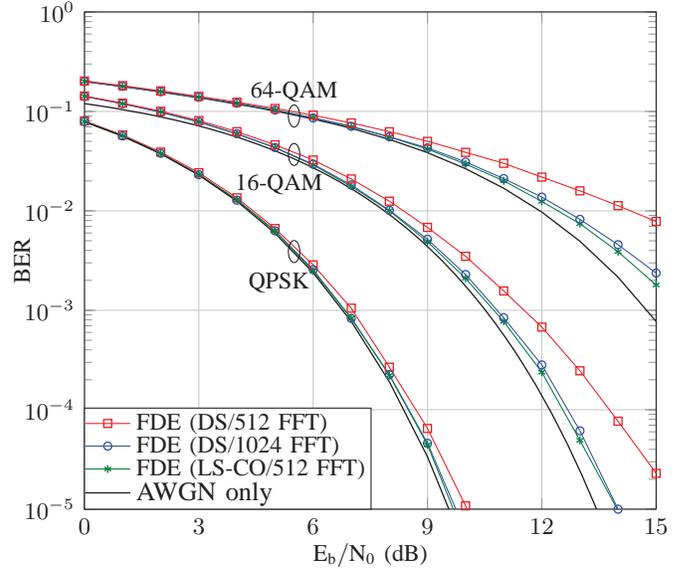


Figure 12. BER performance of the FDE with LS-CO and DS filter as the component of overlap-save method for 1000 km fiber length and QPSK, 16-QAM, and 64-QAM modulations with 4-bit filter coefficient word length.

of the overlap-save method. The fiber length is 1000 km. We show results for the LS-CO filter with  $0.6N_{\text{max}}$  (241 taps), and the DS filter with  $0.6N_{\text{max}}$  (241 taps) and  $N_{\text{max}}$  (403 taps). The shortest possible fast Fourier transform (FFT) size for overlap-save method is selected and the rest of parameters are taken from Table I. We assume perfect FFT/IFFT operations in the overlap-save method and the same quantization method explained in Section III-B is applied for quantizing the frequency response of the filter after FFT operation using 4-bit word length. As can be seen, the FDE with LS-CO component with an FFT size of 512 has better performance than the FDE with DS component with an FFT size of 512 and 1024. Furthermore, the performance gain improves with the modulation order.

## V. CONCLUSION

We have presented a novel finite impulse response filter for chromatic dispersion compensation which is robust to filter tap quantization errors due to finite-precision arithmetic. We observed that, for finite precision, a higher out-of-band gain leads to a higher quantization error. Therefore, the proposed filter, presented in closed form, is designed based on confining the out-of-band gain (in the constrained least-squares sense). Simulation results show that our proposed filter has less sensitivity to finite precision leading to better performance compared with other filters proposed in the literature. The performance gain increases for higher modulation order and lower number of bits used to represent the filter taps. We also show that the nominal number of filter taps ( $N_{\text{max}}$ ) can be reduced up to 40% with little loss in performance. However, a further reduction significantly degrades the performance.

The proposed filter can also be used as the component of the overlap-save method for the FDE in scenarios where the accumulated CD is large, such as for long links.

APPENDIX A  
PROOF OF THEOREM 1

The Lagrangian function of (10) can be expressed as

$$L(\lambda) = \xi_s - \lambda(\xi_o - \xi_{o,\max}). \quad (18)$$

According to (2) and (7)–(9), by taking the derivative of the Lagrangian function with respect to the complex conjugate of the filter tap coefficients, a system of  $N$  equations with  $N$  variables is found where the  $n$ th equation is

$$\begin{aligned} \frac{\partial}{\partial h_n^H} L(\lambda) &= \frac{1}{\Omega_2 - \Omega_1} \int_{\Omega_1}^{\Omega_2} \left( h_n + \sum_{\substack{m=-\frac{N-1}{2} \\ m \neq n}}^{\frac{N-1}{2}} h_m e^{j(n-m)\omega T} \right. \\ &\quad \left. - e^{jM(\omega T)^2 + jn\omega T} \right) d(\omega T) - \frac{\lambda}{2\pi + \Omega_1 - \Omega_2} \\ &\quad \left( \int_{-\pi}^{\Omega_1} \left( h_n + \sum_{\substack{m=-\frac{N-1}{2} \\ m \neq n}}^{\frac{N-1}{2}} h_m e^{j(n-m)\omega T} \right) d(\omega T) + \right. \\ &\quad \left. \int_{\Omega_2}^{\pi} \left( h_n + \sum_{\substack{m=-\frac{N-1}{2} \\ m \neq n}}^{\frac{N-1}{2}} h_m e^{j(n-m)\omega T} \right) d(\omega T) \right) = 0. \end{aligned} \quad (19)$$

By ordering the variables and computing the integrals, the system of equations can be rewritten in matrix form as

$$\mathbf{Q}\tilde{\mathbf{h}} = \mathbf{v}, \quad (20)$$

where  $\mathbf{Q}$  and  $\mathbf{v}$  are given in (13)–(14). The filter tap coefficients calculated based on (13) are dependent on the Lagrangian parameter, which can be computed by holding the complementary slackness condition [16]

$$\xi_o = \xi_{o,\max}. \quad (21)$$

Since the FIR filter is an all-pass filter for pass-band frequencies, we can assume that  $|\tilde{H}(e^{j\omega T})|^2 = 1$  for the  $\Omega_s$  band. Hence,  $\xi_o$  can be simplified as

$$\frac{1}{2\pi + \Omega_1 - \Omega_2} \left[ \int_{-\pi}^{\pi} |\tilde{H}(e^{j\omega T})|^2 d(\omega T) - (\Omega_2 - \Omega_1) \right]. \quad (22)$$

Now, using Parseval's theorem for the DFT [14],

$$\sum_{n=-\infty}^{\infty} |\tilde{h}_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\tilde{H}(e^{j\omega T})|^2 d(\omega T), \quad (23)$$

in (22) and plugging the result into (21), yields the following equation,

$$\frac{2\pi\tilde{\mathbf{h}}^H\tilde{\mathbf{h}} + \Omega_1 - \Omega_2}{2\pi + \Omega_1 - \Omega_2} = \xi_{o,\max}. \quad (24)$$

By solving (24), the Lagrangian parameter  $\lambda$  is found. This concludes the proof. ■

REFERENCES

- [1] S. J. Savory, "Digital coherent optical receivers: Algorithms and sub-systems," *IEEE J. Select. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1164–1179, Sep./Oct. 2010.
- [2] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Express*, vol. 16, no. 2, pp. 753–791, Jan. 2008.
- [3] S. Ramachandran, *Fiber based dispersion compensation*. Springer Science & Business Media, 2007.
- [4] M. G. Taylor, "Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments," *IEEE Photon. Technol. Lett.*, vol. 16, no. 2, pp. 674–676, Feb. 2004.
- [5] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–817, Jan. 2008.
- [6] A. Eghbali, H. Johansson, O. Gustafsson, and S. J. Savory, "Optimal least-squares FIR digital filters for compensation of chromatic dispersion in digital coherent optical receivers," *J. Lightw. Technol.*, vol. 32, no. 8, pp. 1449–1456, Apr. 2014.
- [7] B. S. G. Pillai, B. Sedighi, W. Shieh, and R. S. Tucker, "Chromatic dispersion compensation – an energy consumption perspective," in *Proc. Optical Fiber Communication Conf. (OFC)*, Los Angeles, CA, Mar. 2012.
- [8] B. S. G. Pillai, B. Sedighi, K. Guan, N. P. Anthapadmanabhan, W. Shieh, K. J. Hinton, and R. S. Tucker, "End-to-end energy modeling and analysis of long-haul coherent transmission systems," *J. Lightw. Technol.*, vol. 32, no. 18, pp. 3093–3111, Sep. 2014.
- [9] T. Xu, G. Jacobsen, S. Popov, J. Li, E. Vanin, K. Wang, A. T. Friberg, and Y. Zhang, "Chromatic dispersion compensation in coherent transmission system using digital filters," *Opt. Express*, vol. 18, no. 15, pp. 16243–16257, Jul. 2010.
- [10] D. E. Crivelli, M. R. Hueda, H. S. Carrer, M. del Barco, R. R. Lopez, P. Gianni, J. Finochietto, N. Swenson, P. Voois, and O. E. Agazzi, "Architecture of a single-chip 50 Gb/s DP-QPSK/BPSK transceiver with electronic dispersion compensation for coherent optical channels," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 4, pp. 1012–1025, Apr. 2014.
- [11] A. Sheikh, C. Fougstedt, P. Johannisson, A. Graell i Amat, and P. Larsson-Edefors, "Dispersion compensation filter design optimized for robustness and power efficiency," in *Proc. Signal Processing in Photonic Communications (SPPCom)*, Boston, MA, Jul. 2015, p. SpT3D.2.
- [12] C. Fougstedt, A. Sheikh, P. Johannisson, A. Graell i Amat, and P. Larsson-Edefors, "Power-efficient time-domain dispersion compensation using optimized FIR filter implementation," in *Proc. Signal Processing in Photonic Communications (SPPCom)*, Boston, MA, Jul. 2015, p. SpT3D.3.
- [13] J. C. Geyer, C. R. Fludger, T. Duthel, C. Schullien, and B. Schmauss, "Efficient frequency domain chromatic dispersion compensation in a coherent polmux qpsk-receiver," in *Proc. Optical Fiber Communication Conf. (OFC)*, 2010, p. OWV5.
- [14] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-time signal processing*. Prentice-hall Englewood Cliffs, 1989.
- [15] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2008.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [17] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Washington, DC, 1964.