Back propagation control of model-based multi-layer adaptive filters for optical communication systems

Manabu Arikawa

Department of Advanced Mathematical Sciences
Graduate School of Informatics
Kyoto University

June 2023

Abstract

Optical fiber communication systems require digital signal processing for efficient and precise impairment compensation. Imperfection in transmitter and receiver devices is becoming one of the limiting factors as the symbol rate increases and complicated advanced modulation formats with high spectral efficiency are adopted. The characteristics of impairments due to device imperfection are usually unknown, and an adaptive filter, in which the filter coefficients are controlled on the basis of the filter output so that it approaches the desired feature, is commonly used to compensate for such unknown impairments. Meanwhile, conventional digital signal processing consists of multiple filter blocks to compensate for various impairments that occur in an optical fiber communication system in a block-wise manner. In this architecture, an adaptive filter can be applied to the last filter block since the outputs, except for the one of the last block, include signal distortion due to impairments that should be compensated in other blocks, and thus the objective function to be minimized for an adaptive filter is hard to construct. This fact makes efficient compensation of transmitter and receiver imperfection difficult.

In this thesis, we investigate an adaptive multi-layer filter approach, in which the filter structures in layers are determined by physical models of impairments that occur in an optical fiber communication system by considering the mutual non-commutativity of impairments. By applying the data-driven approach found in the learning of deep neural networks to these unfolded multi-layer filters, the filter coefficients in all the layers can be controlled by gradient descent and back propagation from the last layer's outputs. This data-driven control of the model-based multi-layer filters provides efficient and precise impairment compensation even for transmitter and receiver imperfection.

This thesis is organized as follows. Chapter 2 reviews the fundamentals of finite impulse response filters and the mathematical models of impairments in optical fiber communication systems.

Chapter 3 investigates an adaptive multi-layer filter architecture to efficiently compensate for linear IQ impairments that occur in a transmitter and receiver, considering the order of impairments and their mutual non-commutativity. The adaptive filter coefficient control for multiple cascaded filters with stochastic gradient descent and back propagation is addressed. The performance of impairment compensation for IQ skew in a transmitter and receiver by the adaptive multi-layer filters is evaluated in a numerical simulation and a transmission experiment.

The filter coefficients in each layer in the adaptive multi-layer filter architecture after the convergence of the adaptive control should include information on compensated impairments that are unknown before compensation if the adap-

tive multi-layer filter architecture works appropriately. In Chapter 4, the filter coefficients of the adaptive multi-layer filters are utilized for simultaneous and individual monitoring of transmitter and receiver impairments. The model relating IQ impairments to the widely linear filter response to compensate for them by considering the order in which IQ skew, IQ imbalance, and IQ phase deviation occur for the transmitter and receiver is derived.

Linear IQ impairments that occur in a transmitter and receiver are treated in Chapters 3 and 4, while imperfection in transmitter and receiver devices can also cause nonlinear impairments. Chapter 5 extends the adaptive multi-layer filter architecture to compensate for Tx and Rx nonlinearity by incorporating Volterra filter layers.

Although the adaptive multi-layer filter architecture can compensate for impairments that occur in an optical fiber communication system, including transmitter and receiver impairments, efficiently and precisely, in other words, forward propagation is efficient compared to corresponding one large adaptive filter, the computational complexity for back propagation is still large, as the one for deep neural networks. It can be problematic, especially when the adaptive multi-layer filter architecture is applied to an ultra-long-haul transmission since the temporal spread due to chromatic dispersion becomes large. In Chapter 6, the mutual non-commutativity of SL and WL filters is revisited, and an adaptive multi-layer filter architecture to mitigate the computational complexity of back propagation for applying it to an ultra-long-haul optical fiber transmission is proposed.

Back propagation control of model-based multi-layer filters can be widely applicable if the filter models to be considered are differentiable. In Chapter 7, optimization by SGD and back propagation of multiple layers is applied to a filter that includes an internal sampling rate conversion. A frequency-domain adaptive filter with fractional oversampling controlled by SGD and back propagation is investigated for efficient adaptive MIMO equalization with a large temporal spread.

Chapter 8 gives the conclusion and outlook of this work.

Acknowledgment

I would like to express my respect and sincere gratitude to Prof. Kazunori Hayashi for his guidance throughout this work. His deep insight into the problem has improved this research. His constructive comments have been a source of my inspiration.

A part of this work has been done in my workplace. I would like to express my sincere appreciation to Emmanuel Le Taillandier de Gabory for his support and permission to commit to this doctoral work while my working. His discussion and expertise in optical communication systems helped me. I am grateful to Kohei Hosokawa for his support in working.

I would like to thank Masaki Sato for his experimental support of high-frequency electrical components.

I would like to thank Prof. Tadashi Wadayama for his fruitful presentation on the application of deep learning to communication systems, which has inspired the idea of this work.

I am grateful to Dr. Norifumi Kamiya for his advice on FEC schemes and discussion on learning models.

I would also like to acknowledge many others who have supported this work.

Finally, I would like to express my gratitude to my parents and family for the support they gave me throughout my life.

Contents

Chapter 1

Introduction

Introducing coherent detection and digital signal processing (DSP), referred to as digital coherent [1–3], into optical fiber communication systems has opened the way to flexible compensation of various impairments that occur in optical fiber communication systems including carrier phase and frequency offset [4], chromatic dispersion (CD) [5], polarization mode dispersion [6], and nonlinearity through fiber propagation [7]. Not just effects that occur in optical fiber propagation, impairments due to the imperfection of optical or electrical components in a transmitter and receiver are also dealt with DSP [8, 9]. As the adoption of advanced modulation formats with a high spectral efficiency [10–12] as well as a higher symbol rate [13–15] have proceeded, precise compensation of these various effects in optical fiber communication systems is becoming crucial. Since optical fiber communication systems are usually for large-capacity transmission, DSP for them consequently is required to operates at high throughput. Therefore, high efficiency in terms of computational complexity, as well as precise compensation, is necessary for DSP.

DSP for impairment compensation of optical fiber communication systems is regarded as a filter in a broad sense. Filters for impairment compensation can be categorized as deterministic and adaptive, ones in which the model and parameters of impairments to be compensated are fully known; and ones in which the filter response is adjusted by using filter inputs and outputs. Chromatic dispersion compensation is usually performed by using an all-pass filter with the inverse response of accumulated CD in a transmission line on the bases of its theoretical model [5,16], and it is an example of the former. Compensation of nonlinear impairment that occurs in optical fiber propagation of a signal with a conventional split-step back propagation method is also deterministic [7,17–19]. A well-known example of a later adaptive filter in optical fiber communication systems is one for compensation of polarization state variation and polarization mode dispersion in optical fiber propagation [5,6,20,21]. Compensation of phase and frequency offset between a carrier and a local oscillator, also referred to as carrier recovery [4,22–24], is broadly adaptive since it exploits information from input signals.

Impairments that occur in an optical communication system have different features, such as linear or nonlinear, single-input single-output process or multi-input or multi-output (MIMO), fully known or unknown. Conventional DSP for optical fiber communication systems deals with these multiple impairments in a block-wise manner; that is, multiple impairments are compensated individually in different ways by deterministic or nondeterministic filters according to the characteristics of each impairment. One reason to adopt this block-wise architecture is efficiency in terms of computational complexity. A filter that can compensate for all the relevant impairments in a lump must be redundantly large, with full cross-terms and nonlinear terms having a large temporal spread.

When we look to other communities, machine learning approaches, in which the whole process is learned by exploiting intrinsic relations of given data, have widely been used and succeeded. Deep learning based on neural networks optimized by gradient descent with back propagation has been a major trend. It has greatly succeeded in pattern recognition and natural language processing [25]. Machine learning approaches are regarded as versatile since they do not rely on elaborate designs of factorized individual functions that comprise the interested process. In optical fiber communication communities, machine learning-based approaches have also been investigated, for example, for optical performance monitoring from received waveforms [26], design of optical Raman amplifiers [27], end-to-end optimization of a signal format and receiver-side processing [28–30], and compensation of nonlinearity that occurs in optical fiber propagation and transmitter (Tx) and receiver (Rx) devices [31–33]. The application of machine learning in optical fiber communication systems has attracted more attention.

The great success of machine learning approaches in a broad range of fields is based on two features. The one is that a neural network, or multi-layer perceptron, can provide a precise approximated representation for any complicated function to solve a problem [34]. The other is that the desired process can be acquired by learning a parametrized function with given data without explicitly designing the function. The latter is relied on the optimization of parameters by gradient descent and gradient calculation based on the chain rule of derivatives, which is referred to as back propagation [35]. However, machine learning approaches, especially with neural networks and back propagation, have not widely and practically succeeded in impairment compensation for optical fiber communication systems. One of the reasons for that is the computational complexity. Although a neural network has provided satisfactory performance for problems in a wide range, its computational complexity can be problematic for optical fiber communication systems in which a high throughput DSP is usually required. In contrast, it has been pointed out that the approach of back propagation can be applied not only to neural networks but also to any differentiable parametrized function, which is utilized in today's machine learning platforms as automatic differentiation [36]. An example of optimization with back propagation in communication systems can be seen in the optimization of an iterative process by unfolding it and embedding trainable parameters, referred to as deep unfolding [37–39]. In this thesis, we focus on the application of back-propagation-based and data-driven approaches in impairment compensation for optical fiber communication systems.

A conventional adaptive filter for impairment compensation can already be categorized in data-driven approaches. Several algorithms for adaptive control of the filter coefficient are known [5, 40], which are based on gradient descent, though a conventional adaptive filter uses its direct inputs and outputs of the filter. Meanwhile, in conventional DSP of multiple filter blocks for optical fiber communication systems, an adaptive filter can be applied to the last filter block since the outputs, except for the one of the last block, include signal distortion due to impairments that should be compensated in other blocks, and thus the objective function to be minimized for an adaptive filter is hard to construct. Thus, the adaptive control of a conventional adaptive filter is an optimization within a block. Whereas, when outputs of each block are described as a differentiable function in terms of its internal parameters and inputs, optimization with back propagation can be applied over multiple blocks, and it opens a way to the optimization of an entire impairment compensation. Given the above background, we take advantage of back propagation in impairment compensation for optical fiber communication systems to expand and improve a conventional block-wise architecture for impairment compensation with reasonable computational complexity. A filter to compensate for all the relevant impairments that occur in an optical fiber communication system is divided into blocks, i.e., multi-layers, that have different features, such as linear or nonlinear and single-input single-output or MIMO, to compensate for each of the interested impairments. The filter coefficients of all layers are adaptively controlled by using gradient descent and back propagation to minimize the loss function that is constructed from the last layer outputs. This approach is a kind of unfolding of an adaptive filter. This can also be regarded as the use of domain knowledge in applying a machine learning-based approach in optical fiber communication systems. We can introduce several constraints to blocks in layers by considering a physical model of an optical fiber communication system. In addition, reexamining impairment compensation DSP in optical fiber communication systems from the viewpoint of machine learning, which has wide-range applications, should provide further extensibility in incorporating technologies in other fields in the future.

Chapter 2

Fundamentals of filters and optical fiber communication systems

Most impairments that occur in optical fiber communication systems and are also practically subject to compensation are characterized by their impulse response in the time domain or the transfer function in the frequency domain, which stems from the theory of linear time-invariant systems. This chapter reviews the fundamentals of finite impulse response (FIR) filters. Then, we also review the mathematical models of impairments in optical fiber communication systems. In DSP for communication systems, impairments and filters for compensation of them are usually described by models with a discrete-time, though we deal with models with a continuous time here. Signals and impairments have their physical entities in real communication systems.

In this thesis, we use the following notations; $\mathbb{R}$: the set of real numbers, $\mathbb{C}$: the set of complex numbers, $*$: complex conjugate, i: the imaginary unit, T: transpose, †: Hermitian conjugate, $\circ$: Hadamard product, I: an identity matrix with an appropriate size, $\mathcal{F}$: Fourier transform[1], $\mathcal{F}^{-1}$: inverse Fourier transform, $\star$: convolution.

2.1 FIR filters

In communication systems, we often deal with signals that have narrow spectral components around a specific frequency [41]. These signals are referred to as band-pass signals and the frequency is referred to as a carrier frequency. A band-pass signal $x_b(t) \in \mathbb{R}$ at a carrier frequency $\omega_0/(2\pi)$ is represented as

$$x_b(t) = \mathrm{Re}[x(t)e^{i\omega_0 t}], \tag{2.1}$$

where $x(t) \in \mathbb{C}$ is the complex envelope of the band-pass signal. t is the continuous time. The band-pass signal $x_b(t)$ has its spectral component around the angular frequencies of ω_0 and $-\omega_0$, whereas the complex envelop $x(t)$ has a

---

[1] $\mathcal{F}[x](\omega) = X(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t) \exp(i\omega t) dt$

spectral component around the zero frequency. Thus, the complex envelope $x(t)$ is also called the low-pass equivalent. The real and imaginary parts of $x(t)$ are referred to as in-phase (I) and quadrature (Q) components, which are denoted by $x_{\mathrm{I}}(t), x_{\mathrm{Q}}(t) \in \mathbb{R}$. Thus,

$$x(t) = x_{\mathrm{I}}(t) + i x_{\mathrm{Q}}(t). \tag{2.2}$$

We refer to a low-pass equivalent signal just as a signal here.

An effect on a signal or its compensation is modeled by a filter. Since we deal with complex-valued signals, a filter is usually modeled with a complex-valued response and complex-valued input and output. However, this filter model imposes a non-negligible restriction to a response. That is, a complex-valued filter with a complex-valued input and output cannot handle IQ components independently [42]. Some impairments that occur in optical communication systems, especially in Tx and Rx devices, for example, the timing misalignment between IQ components (IQ skew), the difference of average amplitudes of IQ components (IQ imbalance), and the orthogonal phase deviation between IQ components from $\pi/2$ (IQ phase deviation), have different responses for IQ components though they are linear. They cannot be modeled by a complex-valued filter with a complex-valued input and output signal. In this literature, these filters are referred to as strictly linear (SL). A real-valued MIMO filter with real-valued inputs and outputs of IQ components is required to model a linear process with different responses for IQ components. As we will see later, a real-valued MIMO filter with real-valued inputs and outputs is equivalent to a complex-valued response filter whose inputs are a complex-valued signal and its complex conjugate, which is referred to as widely linear.

2.1.1 SL filters

We first deal with SL filters. The input and output of an SL filter with single input and single output $x(t), y(t) \in \mathbb{C}$ are related with convolution as

$$y(t) = \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau, \tag{2.3}$$

where $h(t) \in \mathbb{C}$ is an impulse response of the filter. In the case of a SL filter with K inputs and K outputs, the inputs $x_1(t), \ldots, x_K(t) \in \mathbb{C}$ and the outputs $y_1(t), \ldots, y_K(t) \in \mathbb{C}$ are related as

$$y_i(t) = \sum_{j=1}^{K} \int_{-\infty}^{\infty} h_{ij}(\tau) x_j(t-\tau) d\tau. \tag{2.4}$$

The frequency-domain representation of Eq. (2.4) is given as

$$Y_i(\omega) = \sum_{j=1}^{K} H_{ij}(\omega) X_j(\omega). \tag{2.5}$$

When we consider the input and output vectors of the SL filter with K inputs and K outputs, they are related as

$$\begin{pmatrix} Y_1(\omega) \\ \vdots \\ Y_K(\omega) \end{pmatrix} = \begin{pmatrix} H_{11}(\omega) & \cdots & H_{1K}(\omega) \\ \vdots & \ddots & \vdots \\ H_{K1}(\omega) & \cdots & H_{KK}(\omega) \end{pmatrix} \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_K(\omega) \end{pmatrix}, \tag{2.6}$$

and the matrix

$$H^{\mathbb{C}}(\omega) = \begin{pmatrix} H_{11}(\omega) & \cdots & H_{1K}(\omega) \\ \vdots & \ddots & \vdots \\ H_{K1}(\omega) & \cdots & H_{KK}(\omega) \end{pmatrix} \tag{2.7}$$

includes all the information of the response of this SL filter.

On the other hand, the complex-valued signal $x(t)$ can be represented by its IQ components $x_{\mathrm{I}}(t), x_{\mathrm{Q}}(t) \in \mathbb{R}$, as $x(t) = x_{\mathrm{I}}(t) + ix_{\mathrm{Q}}(t)$. When the filter response $h(t)$ is represented by its real and imaginary parts as $h(t) = h_{\mathrm{I}}(t) + ih_{\mathrm{Q}}(t)$, the input-output relation of a SL filter of Eq. (2.3) is rewritten as

$$\begin{pmatrix} y_{\mathrm{I}}(t) \\ y_{\mathrm{Q}}(t) \end{pmatrix} = \int_{-\infty}^{\infty} \begin{pmatrix} h_{\mathrm{I}}(\tau) & -h_{\mathrm{Q}}(\tau) \\ h_{\mathrm{Q}}(\tau) & h_{\mathrm{I}}(\tau) \end{pmatrix} \begin{pmatrix} x_{\mathrm{I}}(t-\tau) \\ x_{\mathrm{Q}}(t-\tau) \end{pmatrix} d\tau, \tag{2.8}$$

which implies that SL filters cannot handle the IQ components independently. We refer to a representation of Eqs. (2.3) or (2.4) as a complex-valued phasor representation and a representation of Eq. (2.8) as a real-valued IQ representation.

2.1.2 WL filters and their equivalent IQ representations

Equation (2.8) corresponds to a real-valued 2×2 MIMO filter, though a strict constraint is imposed on the real-valued response matrix to be an SL filter. In contrast, a general real-valued 2×2 MIMO filter is written as

$$\begin{pmatrix} y_{\mathrm{I}}(t) \\ y_{\mathrm{Q}}(t) \end{pmatrix} = \int_{-\infty}^{\infty} \begin{pmatrix} h_{\mathrm{II}}(\tau) & h_{\mathrm{IQ}}(\tau) \\ h_{\mathrm{QI}}(\tau) & h_{\mathrm{QQ}}(\tau) \end{pmatrix} \begin{pmatrix} x_{\mathrm{I}}(t-\tau) \\ x_{\mathrm{Q}}(t-\tau) \end{pmatrix} d\tau. \tag{2.9}$$

The representation of Eq. (2.9) can handle the different responses independent of IQ components, such as IQ skew and IQ crosstalk. The frequency-domain representation of Eq. (2.9) is

$$\begin{pmatrix} Y_{\mathrm{I}}(\omega) \\ Y_{\mathrm{Q}}(\omega) \end{pmatrix} = \begin{pmatrix} H_{\mathrm{II}}(\omega) & H_{\mathrm{IQ}}(\omega) \\ H_{\mathrm{QI}}(\omega) & H_{\mathrm{QQ}}(\omega) \end{pmatrix} \begin{pmatrix} X_{\mathrm{I}}(\omega) \\ X_{\mathrm{Q}}(\omega) \end{pmatrix}, \tag{2.10}$$

where

$$H^{\mathbb{R}}(\omega) = \begin{pmatrix} H_{\mathrm{II}}(\omega) & H_{\mathrm{IQ}}(\omega) \\ H_{\mathrm{QI}}(\omega) & H_{\mathrm{QQ}}(\omega) \end{pmatrix} \tag{2.11}$$

is the frequency-domain response of a real-valued 2×2 MIMO filter.

A WL filter corresponds to a complex-valued phasor representation of Eq. (2.9). A 2×1 WL filter[2], in which the complex-valued inputs are $x(t)$ and its complex conjugate $x^*(t)$, is represented by

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau + \int_{-\infty}^{\infty} h_*(\tau)x^*(t-\tau)d\tau, \tag{2.12}$$

where $h(t), h_*(t) \in \mathbb{C}$ are complex-valued filter impulse responses. Equation (2.12) can be rewritten as

$$y(t) = \int_{-\infty}^{\infty} (h(\tau)\, h_*(\tau)) \begin{pmatrix} x(t-\tau) \\ x^*(t-\tau) \end{pmatrix} d\tau. \tag{2.13}$$

---

[2] We denote the size of a filter by the size of the response matrix.

Furthermore, it can be written when we use the signal vector $(y(t),\, y^*(t))^{\mathrm{T}}$ augmented with the complex conjugate as

$$\begin{pmatrix} y(t) \\ y^*(t) \end{pmatrix} = \int \begin{pmatrix} h(\tau) & h_*(\tau) \\ h_*^*(\tau) & h^*(\tau) \end{pmatrix} \begin{pmatrix} x(t-\tau) \\ x^*(t-\tau) \end{pmatrix} d\tau. \tag{2.14}$$

The augmented signal vector $(y(t),\, y^*(t))^{\mathrm{T}}$ with the complex conjugate is related to the signal vector in the real-valued IQ representation $(y_{\mathrm{I}}(t),\, y_{\mathrm{Q}}(t))^{\mathrm{T}}$ as

$$\begin{pmatrix} y(t) \\ y^*(t) \end{pmatrix} = T \begin{pmatrix} y_{\mathrm{I}}(t) \\ y_{\mathrm{Q}}(t) \end{pmatrix}, \tag{2.15}$$

where

$$T = \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}, \tag{2.16}$$

and T satisfies

$$T^\dagger T = T T^\dagger = 2I. \tag{2.17}$$

Multiplying T from the left of Eq. (2.9), we obtain

$$\begin{pmatrix} h(\tau) & h_*(\tau) \\ h_*^*(\tau) & h^*(\tau) \end{pmatrix} = \frac{1}{2} T \begin{pmatrix} h_{\mathrm{II}}(\tau) & h_{\mathrm{IQ}}(\tau) \\ h_{\mathrm{QI}}(\tau) & h_{\mathrm{QQ}}(\tau) \end{pmatrix} T^\dagger, \tag{2.18}$$

or

$$\begin{pmatrix} h_{\mathrm{II}}(\tau) & h_{\mathrm{IQ}}(\tau) \\ h_{\mathrm{QI}}(\tau) & h_{\mathrm{QQ}}(\tau) \end{pmatrix} = \frac{1}{2} T^\dagger \begin{pmatrix} h(\tau) & h_*(\tau) \\ h_*^*(\tau) & h^*(\tau) \end{pmatrix} T. \tag{2.19}$$

Therefore, a 2×1 WL filter and a real-valued 2×2 MIMO filter with the IQ representation are equivalent. In the frequency domain, Eq. (2.19) corresponds to

$$\begin{pmatrix} H_{\mathrm{II}}(\omega) & H_{\mathrm{IQ}}(\omega) \\ H_{\mathrm{QI}}(\omega) & H_{\mathrm{QQ}}(\omega) \end{pmatrix} = \frac{1}{2} T^\dagger \begin{pmatrix} H(\omega) & H_*(\omega) \\ H_*^*(-\omega) & H^*(-\omega) \end{pmatrix} T. \tag{2.20}$$

In the case of a $2K \times K$ MIMO WL filter, the input-output relation is

$$y_i(t) = \sum_{j=1}^{K} \left(\int_{-\infty}^{\infty} h_{ij}(\tau) x_j(t-\tau) d\tau + \int_{-\infty}^{\infty} h_{*ij}(\tau) x_j^*(t-\tau) d\tau \right). \tag{2.21}$$

When we consider a augmented signal $\underline{z}$ with the complex conjugate for $z = x$ or y as

$$\underline{z}_i = \begin{cases} z_i & (i = 1, \ldots, K) \\ z_{i-K}^* & (i = K+1, \ldots, 2K) \end{cases} \tag{2.22}$$

whose length is $2K$, and also consider $\hat{h}_{ij}$ as

$$\hat{h}_{ij} = \begin{cases} h_{ij} & (i, j = 1, \ldots, K) \\ h_{*i(j-K)} & (i = 1, \ldots, K,\ j = K+1, \ldots, 2K) \end{cases} \tag{2.23}$$

whose size is $K \times 2K$, Eq. (2.21) becomes

$$y_i(t) = \sum_{j=1}^{2K} \int_{-\infty}^{\infty} \hat{h}_{ij}(\tau) \underline{x}_j(t-\tau) d\tau. \tag{2.24}$$

Moreover, considering

$$\underline{h}_{ij} = \begin{cases} h_{ij} & (i,j=1,\ldots,K) \\ h_{*i(j-K)} & (i=1,\ldots,K,\ j=K+1,\ldots,2K) \\ h^*_{*(i-K)j} & (i=K+1,\ldots,2K,\ j=1,\ldots,K) \\ h^*_{(i-K)(j-K)} & (i,j=K+1,\ldots 2K) \end{cases} \tag{2.25}$$

whose size is $2K \times 2K$, the augmented output $\underline{y}$ is given by

$$\underline{y}_i(t) = \sum_{j=1}^{2K} \int_{-\infty}^{\infty} \underline{h}_{ij}(\tau) \underline{x}_j(t-\tau) d\tau, \tag{2.26}$$

which results in

$$\begin{pmatrix} \underline{y}_1(t) \\ \vdots \\ \underline{y}_{2K}(t) \end{pmatrix} = \int_{-\infty}^{\infty} \begin{pmatrix} \underline{h}_{11}(\tau) & \cdots & \underline{h}_{1(2K)}(\tau) \\ \vdots & \ddots & \vdots \\ \underline{h}_{(2K)1}(\tau) & \cdots & \underline{h}_{(2K)(2K)}(\tau) \end{pmatrix} \begin{pmatrix} \underline{x}_1(t-\tau) \\ \vdots \\ \underline{x}_{2K}(t-\tau) \end{pmatrix} d\tau. \tag{2.27}$$

2.1.3 Adaptive filters

A filter whose response is autonomously controlled by using its input and output with a certain criterion is referred to as an adaptive filter. Adaptive filters are widely used in systems including unknown responses for inverse modeling and interference cancellation [40]. In optical fiber communication systems, an adaptive filter is used to compensate for polarization state variation and polarization mode dispersion, which are stochastic processes and time-varying, over an optical fiber transmission line [5, 6].

An adaptive filter for impairment compensation in optical fiber communication systems is usually implemented in the digital domain[3]. An adaptive linear filter is modeled as shown in Fig. 2.1. The input signal x of the adaptive filter suffers some impairments to be compensated, and the objective of the adaptive filter is to obtain the output signal y in the sense of minimizing a certain error from a desired property by adjusting the filter response h. We consider a SISO FIR filter[4] with a complex-valued response with a length of M, as $\boldsymbol{h} = (h[0], \ldots, h[M-1])^{\mathrm{T}}$. The input and output signals $x, y \in \mathbb{C}$ at a timing integer k are related with a convolution, as

$$y[k] = \sum_{m=0}^{M-1} h[m]x[k-m] = \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}, \tag{2.28}$$

where $\boldsymbol{x} = (x[k], \ldots, x[k-M+1])^{\mathrm{T}}$. The filter coefficients $\boldsymbol{h}$ are updated by an algorithm to minimize the loss function consisting of an error of the output from a desired property. Constant modulus algorithm (CMA) and data-aided or decision-directed LMS are widely used in optical fiber communication systems. These algorithms are based on gradient descent. Since a generalized loss is hard to treat, as in the case of learning neural networks, stochastic gradient descent is adopted.

---

[3] Analog systems can be implemented in the digital domain with a sufficiently high sampling rate [43].

[4] An extension to a MIMO filter is straightforward.

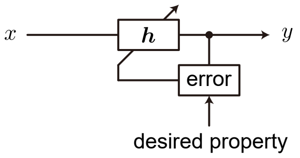


Figure 2.1. Model of adaptive filter for impairment compensation

Wirtinger derivatives

In communication systems, complex-valued signals and filter coefficients are treated. A loss function should be real-valued to be minimized. Gradient descent of filter coefficients in terms of the loss function is regarded as a simultaneous update of real and imaginary parts of filter coefficients with partial derivatives. In this case, Wirtinger derivatives give a good perspective.

When f is a function of independent real-valued variables u and v via a complex-valued variable $z = u + iv$, the total derivative df is given by using partial derivatives $\partial f/\partial u, \partial f/\partial v$, as

$$df = \frac{\partial f}{\partial u} du + \frac{\partial f}{\partial v} dv. \tag{2.29}$$

Small amounts of du, dv are related to dz and dz^* as

$$dz = du + idv, \tag{2.30}$$

$$dz^* = du - idv. \tag{2.31}$$

The total derivative of Eq. (2.29) can be rewritten as

$$df = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial z} dz^*, \tag{2.32}$$

where

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial u} - i\frac{\partial f}{\partial v}\right), \tag{2.33}$$

$$\frac{\partial f}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f}{\partial u} + i\frac{\partial f}{\partial v}\right) \tag{2.34}$$

are Wirtinger derivatives[5].

We consider the update of the real-valued variables u and v with gradient descent regarding f. When u, v are updated by

$$u \to u - \alpha \frac{\partial f}{\partial u}, \tag{2.35}$$

---

[5] The Cauchy–Riemann equations, which are necessary and sufficient conditions for complex differentiable, or holomorphic, corresponds to $\partial f/\partial z^* = 0$.

$$v \to v - \alpha \frac{\partial f}{\partial v}, \tag{2.36}$$

where α is the step size, z is updated by

$$z \to z - 2\alpha \frac{\partial f}{\partial z^*}, \tag{2.37}$$

according to Eqs. (2.33) and (2.34). The above corresponds to the update rule of a complex-valued variable by gradient descent.

Constant modulus algorithm

CMA is an algorithm for the adaptive control of coefficients of an adaptive filter. The loss function of CMA is the squared magnitude of the error of the output from a constant amplitude, which is especially suitable for modulation formats having a constant amplitude, such as QPSK. The instantaneous loss function ϕ is

$$\phi = (r^2 - |y[k]|^2)^2, \tag{2.38}$$

where r is the desired amplitude. According to the update rule by gradient descent of Eq. (2.37), the filter coefficients are updated by

$$\boldsymbol{h} \to \boldsymbol{h} + 4\alpha e \boldsymbol{x}^* y[k], \tag{2.39}$$

where $e = r^2 - |y[k]|^2$.

Least mean square algorithm

The LMS algorithm uses an error of the output of the filter from the desired output symbol to consist of a loss function. The data-aided LMS utilizes pre-known desired symbols, which are transmitted as a training pattern. The decision-directed LMS uses the feature that the transmitted symbols are the elements of the set of the modulation format, and the symbols after decision are used as desired symbols. The instantaneous loss function of the LMS algorithm is

$$\phi = |d[k] - y[k]|^2, \tag{2.40}$$

where $d[k]$ is a known training symbols in the case of the data-aided LMS or a decision output of $y[k]$ in the case of the decision-directed LMS. The filter coefficients are updated by

$$\boldsymbol{h} \to \boldsymbol{h} + 2\alpha e \boldsymbol{x}^*, \tag{2.41}$$

where $e = d[k] - y[k]$.

2.2 Impairments in optical fiber communication systems

In this study, we consider an optical fiber communication system with coherent detection whose schematic diagram is shown in Fig. 2.2. Digital data is converted to continuous electric signals with a digital-to-analog converter (DAC) with a certain encoding. They are mapped to IQ components of an optical signal having two orthogonal polarization modes, i.e., PDM, by modulating a continuous-wave (CW) source from a laser diode (LD). An optical signal is multiplexed to WDM signals and transmitted through optical fiber spans consisting of an SMF and optical repeater amplifier that compensates for propagation loss of a span. An optical signal after optical fiber transmission is demultiplexed to a single WDM channel, and IQ components of two polarization modes are received with coherent detection. In optical fiber communication systems with coherent detection, received signals of IQ components are usually sampled by an analog-to-digital converter, and demodulation and decoding including impairment compensation are performed in the digital domain.

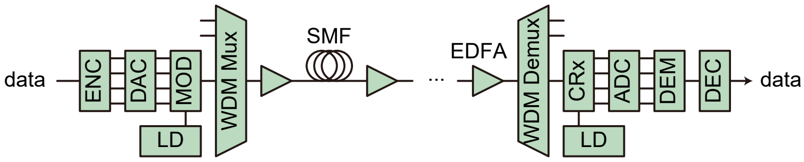


Figure 2.2. Schematic diagram of a optical fiber communication system with WDM and coherent detection. ENC: encoding, DAC: digital-to-analog converter, MOD: modulation, LD: laser diode, SMF: single-mode fiber, EDFA: erbium-doped fiber amplifier, CRx: coherent receiver, ADC: analog-to-digital converter, DEM: demodulation, DEC: decoding

Several impairments occur in the optical fiber communication system. Some impairments occur in the optical domain, that is, in optical fiber propagation. Others occur in the electrical domain, that is, in the Tx or Rx electric devices. Although some impairments are nonlinear, most dominant effects are modeled as linear since the systems operate around the linear region to obtain a good performance. A response of a linear impairment in the optical fiber communication systems is represented by an SL or WL filter. We refer to impairment and a filter that compensates for it together as a process.

Linear impairments that occur in the optical domain, including CD, PMD, polarization state variation, and the carrier phase/frequency offset due to the frequency difference of the signal carrier and the LO source in the Rx[6][7], are modeled by SL processes since an optical signal is modeled by its complex-valued phasor representation, or lowpass equivalent.

In contrast, linear impairments that occur in the electrical domain[8], including the timing misalignment of IQ components (IQ skew), the mismatch of av-

---

[6] In optical fiber propagation, nonlinear distortion due to the optical Kerr effect is non-negligible. A polarization dependent loss is another impairment in the optical domain, though it is usually tiny in an SMF transmission.

[7] In long-haul SMF transmission systems, amplified spontaneous emission (ASE) from optical repeater amplifiers to compensate for span losses is a dominant limiting factor for the transmission performance. ASE noise is a non-deterministic process and it is modeled by an additive white Gaussian (AWGN) noise. We concern here deterministic effects that can be compensated. The performance of a WDM transmission system over SMF spans, including nonlinear distortion, is well estimated by the GN model [44].

[8] More precisely, linear impairments that occur when IQ components are treated independently.

eraged amplitudes of IQ components (IQ imbalance), and the orthogonalization error of IQ components (IQ phase deviation), affects IQ components differently. Thus, they are modeled by WL processes. We review the physical and mathematical model of each dominant impairment in the optical fiber communication system.

2.2.1 Linear impairments in optical domain

A conventional SMF supports two propagation modes of orthogonal polarizations. Thus, propagation effects over an SMF are generally modeled by a 2×2 process. In the cases of a linear effect, the frequency response becomes a 2×2 matrix[9]. Optical propagation over an SMF towards the z axis with random polarization coupling is modeled by a nonlinear Schrödinger equation, up to the second order, as [46, 47]

$$\frac{\partial}{\partial z}\boldsymbol{u}(z,t) = -A\boldsymbol{u} + iB^{(0)}\boldsymbol{u} - B^{(1)}\frac{\partial}{\partial t}\boldsymbol{u} - \frac{i}{2}B^{(2)}\frac{\partial^2}{\partial t^2}\boldsymbol{u} + i\frac{8}{9}\gamma||\boldsymbol{u}||_2^2\boldsymbol{u}, \tag{2.42}$$

where $\boldsymbol{u} = (u_{\mathrm{X}}, u_{\mathrm{Y}})$ is the optical signal (complex envelop) of two polarizations[10] in the complex-valued phasor representation. $A, B^{(0)}, B^{(1)}, B^{(2)}$ are 2×2 matrices. A corresponds to the propagation loss of the fiber. In the case without PDL, it reduces to $(\alpha/2)I$, where α is the loss coefficient of the electric field, which relates to the conventional fiber propagation loss coefficient α_{dB} in the unit of dB/km as $\alpha_{\mathrm{dB}} = 20\alpha\log_{10} e$. $B^{(0)}$ accounts for coupling between polarization modes. $B^{(1)}$ relates to the inverses of the group velocities of two polarization modes, which especially accounts for PMD. $B^{(2)}$ accounts for CD. The last term of the right-hand side of Eq. (2.42) corresponds to the nonlinear term. $||\cdot||_2$ denotes the L_2 norm. γ relates to the fiber nonlinear refractive index n_2 and the effective area of the core A_{eff} as $\gamma = n_2\omega_0/(cA_{\mathrm{eff}})$ where c and ω_0 are the speed of light in vacuum and the angular reference frequency, respectively. The nonlinear term accounts for various nonlinear effects such as self-phase modulation, cross-phase modulation, and four-wave mixing [46], when the optical signal consists of multiple WDM channels.

Optical signal propagation according to Eq. (2.42) can be calculated in a split-step manner [46]. Namely, we consider a small distance Δz where the linear and nonlinear effects are mutually negligible. Equation (2.42) with only the linear terms without the nonlinear term has an analytical solution in the frequency domain. On the contrary, Eq. (2.42) with only the nonlinear term without the linear terms has an analytical solution in the time domain. In the split-step method, the linear and nonlinear propagation effects in a small step are executed alternatively. This procedure corresponds to ignoring the mutual non-commutativity of the linear and nonlinear effects in a small step.

Although the influence of the nonlinear term on the conventional optical communication systems with WDM is not negligible, the primary sources to (deterministically) change the signal waveform profile are linear terms, especially CD and PMD including the variation of the polarization state. Moreover, these effects can be fully compensated by linear filters in the digital domain at the

---

[9] It corresponds to a Jones matrix [45].

[10] That is, Jones vector.

Rx side[11]. We focus on these linear impairments that occur in optical fiber propagation.

Chromatic dispersion

An optical fiber made of silica is a slightly dispersive media. CD corresponds to the term of $\partial^2 \boldsymbol{u}/\partial t^2$ of Eq. (2.42). In an SMF transmission, the dependence of CD on two polarizations is small and usually neglected. In this case, the matrix $B^{(2)}$ is regarded as $B^{(2)} = \beta_2 I$. The frequency response of CD with the propagation distance z is

$$H_{\mathrm{CD}}^{\mathbb{C}}(\omega) = \begin{pmatrix} H_{\mathrm{CD}}(\omega) & 0 \\ 0 & H_{\mathrm{CD}}(\omega) \end{pmatrix} = H_{\mathrm{CD}}(\omega) I, \tag{2.43}$$

and

$$H_{\mathrm{CD}}(\omega) = \exp\left(-i \frac{D_{\mathrm{CD}} \lambda_0 z}{4\pi c} \omega^2\right), \tag{2.44}$$

where D_{CD} is the dispersion coefficient of the fiber, which is usually designated by the unit of ps/nm/km and is related to β_2 as $D_{\mathrm{CD}} = -2\pi c \beta_2 / \lambda_0^2$. λ_0 is the reference wavelength. CD leads to temporal signal spreading, which can be very large in a long-haul SMF transmission.

Polarization state variation

The two propagation modes of orthogonal polarizations that an SMF support are quasi-degenerate. Coupling between two polarization modes occurs through optical fiber propagation, leading to the variation of the polarization state. Since dissipation of the optical signal including PDL is attributed to the matrix A, the response of the polarization state variation is usually modeled by a unitary matrix R as [50]

$$H_{\mathrm{PS}}^{\mathbb{C}}(\omega) = R. \tag{2.45}$$

Polarization mode dispersion

Although the two polarization modes of an SMF are quasi-degenerate, an SMF shows slight birefringence due to core distortion in the fiber manufacturing process and mechanical bending/pressure in operation. This leads to the difference in the group delays of two polarization modes, more precisely two principal states of polarization. When an optical signal does not align with the principal states of polarization, these different group delays cause temporal signal spreading, referred to as PMD. PMD corresponds to the term of $\partial \boldsymbol{u}/\partial t$ of Eq. (2.42). The degree of birefringence of a fiber depends on its local distortion. We focus on a small fragment Δz_k of a fiber in which $B^{(1)}$ can be regarded as constant. $B^{(1)}$ can be diagonalized by a unitary matrix R as

$$B^{(1)} = R^{-1} \Lambda R, \tag{2.46}$$

---

[11] Many efforts have also been provided to compensation of nonlinear effects in optical fiber propagation [7, 17, 18, 48]. Compensation for nonlinear impairments in the digital domain usually requires a high computational complexity, and its practical implementation is still open. Since linear deterministic impairments can be fully compensated, the fundamental limit in optical fiber communication systems is determined by ASE noise and nonlinear impairments, which results in the well-known nonlinear Shannon limit [49].

and $\Lambda = \mathrm{diag}(\beta_{1+}, \beta_{1-})$. $\beta_{1\pm}$ correspond to the first-order propagation constant of two principal states of polarization. $\boldsymbol{Ru}$ corresponds the principal states of polarization. For optical fiber communication systems, we usually concern only the difference $\Delta\beta_1 = \beta_{1+} - \beta_{1-}$ and $\Lambda = \mathrm{diag}(+\Delta\beta_1/2, -\Delta\beta_1/2)$ since the mean value of $\beta_{1\pm}$ does not contribute to the signal distortion. In this case, the frequency response of PMD of a small segment Δz_k is

$$H^{\mathbb{C}}_{\mathrm{PMD},k}(\omega) = R_k^{-1} \begin{pmatrix} \exp(i\omega\tau_{\mathrm{DGD},k}/2) & 0 \\ 0 & \exp(-i\omega\tau_{\mathrm{DGD},k}/2) \end{pmatrix} R_k, \tag{2.47}$$

where $\tau_{\mathrm{DGD},k} = \Delta\beta_1 \Delta z_k$ is the differential group delay. The overall PMD of a fiber is a multiplication of all small K segments as

$$H^{\mathbb{C}}_{\mathrm{PMD}}(\omega) = H^{\mathbb{C}}_{\mathrm{PMD},1}(\omega) H^{\mathbb{C}}_{\mathrm{PMD},2}(\omega) \cdots H^{\mathbb{C}}_{\mathrm{PMD},K}(\omega). \tag{2.48}$$

The temperature and mechanical stress affect a fiber's local distortions, and thus PMD usually temporally varies in an optical fiber communication environment. The typical speed of variation of PMD in long-haul SMF transmissions is the order of kHz [51][12].

Phase/frequency offset

An optical signal wave is converted to a base-band low-pass equivalent signal by interfering with LO in coherent reception. A CW laser source with a near frequency with a signal laser source operating in free-running, that is, having a slightly different frequency, is used for an LO. This configuration is referred to as intradyne reception [4]. Since modulation of a signal is performed on a signal laser source, the relative phase and frequency offset between the signal and LO laser sources appear as a phase rotation of the low-pass equivalent signal. A phase rotation, especially due to a frequency offset, is modeled by the time-domain multiplication of a time-varying phase factor rather than convolution in the time domain. In conventional coherent optical communication systems with PDM, the two orthogonal polarizations have a common phase and frequency offset since the same laser source is supplied for two polarizations in an optical modulator and a polarization-diversity coherent receiver. Given that the phase and (angular) frequency offset at a time t are θ_{LO} and ω_{LO}[13], the optical signal $y_j(t)$, $(j = X, Y)$ after the phase/frequency offset is written by the signal before the offset $x_j(t)$ as

$$y_j(t) = \exp(i(\theta_{\mathrm{LO}} + \omega_{\mathrm{LO}} t)) x_j(t), \tag{2.49}$$

which indicates the temporal response

$$h_{\mathrm{LO}j}(\tau, t) = \delta(\tau) \exp(i(\theta_{\mathrm{LO}} + \omega_{\mathrm{LO}}(t - \tau))). \tag{2.50}$$

When we can assume the phase temporal variation is small in a certain duration, or we only consider the phase offset θ_{LO}, the frequency response can be regarded as

$$H^{\mathbb{C}}_{\mathrm{LO}}(\omega) = e^{i\theta_{\mathrm{LO}}} I. \tag{2.51}$$

---

[12] It is observed that the variation can be the order of 100 kHz due to lightning [52].

[13] An instantaneous frequency offset is related to the variation of the phase offset as $\omega_{\mathrm{LO}}(t) = \partial\theta_{\mathrm{LO}}(t)/\partial t$.

2.2.2 Linear impairments in electrical domain

Not only in optical fiber propagation, but impairments also occur in electrical/optical devices in a Tx and Rx. Figure 2.3 shows the configuration of a Tx and Rx for a coherent optical fiber communication system. Schematic diagrams for one polarization are shown[14]. In these configurations of a Tx and Rx, I and Q components have independent physical substances, that is, physical (electrical) paths. Thus, impairments on IQ components are modeled by a WL process. Since two polarization modes can be treated independently in a Tx and Rx for PDM, we consider the model of a 2×1 WL process for one polarization mode here[15].

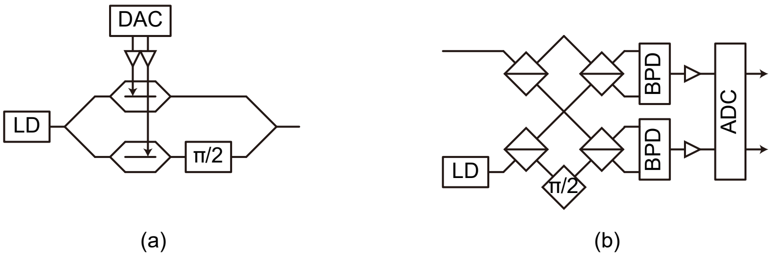


Figure 2.3. Schematic diagram of configuration of (a) Tx and (b) Rx for an optical fiber communication system for one polarization. BPD: balanced photo detector.

IQ skew

IQ skew corresponds to the timing misalignment between the I and Q components. This usually occurs due to the difference in the lengths of electrical paths of IQ components. When the timing delay of the IQ components is $\tau(\omega)$, the frequency response of IQ skew in the IQ representation is

$$H^{\mathbb{R}}_{\text{skew}}(\omega) = \begin{pmatrix} \exp(i\omega\tau(\omega)/2) & 0 \\ 0 & \exp(-i\omega\tau(\omega)/2) \end{pmatrix}, \tag{2.52}$$

which results in the identity matrix when $\tau(\omega) = 0$.

IQ imbalance

IQ imbalance corresponds to the mismatch of the averaged amplitude of the I and Q components, which is usually caused by the non-uniformity of the gain of electrical amplifiers for the I and Q components. The frequency response of

---

[14] We will discuss this configuration of a Tx and Rx in detail in the later chapter.

[15] If we have to treat IQ impairments over two polarization modes, we should consider a 4×2 WL process. This extension is straightforward.

IQ imbalance in the IQ representation is

$$H^{\mathbb{R}}_{\mathrm{imb}}(\omega) = \begin{pmatrix} 1 + a(\omega) & 0 \\ 0 & 1 - a(\omega) \end{pmatrix}, \tag{2.53}$$

where $a(\omega)$ is an indicator of the degree of imbalance[16].

IQ phase deviation

The I and Q components ideally have the phase deviation of $\pi/2$ in the IQ plane, though the imperfection of the devices of an IQ modulator or an optical 90° hybrid causes the phase deviation of the IQ components from $\pi/2$. When the phase deviation from $\pi/2$ is $\phi(\omega)$, the frequency response of IQ phase deviation in the IQ representation is

$$H^{\mathbb{R}}_{\mathrm{phase}}(\omega) = \begin{pmatrix} 1 & \sin\varphi(\omega) \\ 0 & \cos\varphi(\omega) \end{pmatrix}. \tag{2.54}$$

---

[16] For the formalization of the degree of imbalance, there are slightly different styles [53].

Chapter 3

Adaptive multi-layer strictly linear and widely linear filters for compensation of transmitter and receiver impairments

Digital coherent technology in optical fiber communications enables optical fields to be accessed in the digital domain. It providing a flexible way to compensate for various signal distortion. It also opens up possibilities to use advanced modulation formats with high spectral efficiency such as higher order QAM. For example, probabilistic constellation shaping (PCS) based on higher-order QAM formats has attracted much attention because it can provide both fine rate adaptability and high sensitivity approaching the Shannon limit for optical communication systems.

Advanced modulation signals are generally susceptible to various impairments. As a high symbol rate is pursued in optical fiber communication systems up until now, impairments stemming from imperfections of high-speed devices in the Tx and Rx come to have larger impact on the performance. Since the characteristics of these impairments from device imperfections vary slightly due to changes of environment, automatically compensating them with adaptive equalization is desirable. In this chapter, we propose and evaluate an adaptive multi-layer SL&WL filter architecture to efficiently compensate for linear IQ impairments that occur in a Tx and Rx. We first describe the concept of the adaptive multi-layer SL&WL filter architecture and the input-output relations of layers, i.e., forward calculation. We then configure the loss function to be minimized and derive gradient calculation over layers with back propagation. The performance is evaluated in simplified numerical simulation with a 32-Gbaud PDM-QPSK signal and in a transmission experiment of 32-Gbaud

PDM-64QAM. The results show that they can compensate for IQ skew in both Tx and Rx in the presence of CD, polarization rotation, and a frequency offset.

3.1 Background

We consider impairments in optical fiber communication systems and compensation for them in the Rx-side DSP. When we consider a system with coherent detection and PDM, deterministic impairments occur, as shown in Fig. 3.1. On the Tx side, IQ impairments, such as IQ skew and IQ phase deviation in an optical modulator and IQ imbalance in electric driver amplifiers, occur, referred to as Tx impairments. In optical fiber propagation, CD and PMD with polarization state variation occur. As described in the previous chapter, nonlinear distortion due to the optical Kerr effects also occurs together in optical fiber propagation, though we ignore it for simplicity here since linear impairments are more dominant for signal distortion. On the Rx side, the phase and frequency offset occurs in intradyne coherent detection. IQ impairments, such as IQ skew and IQ phase deviation in a coherent receiver and IQ imbalance in electric trans-impedance amplifiers, also occur on the Rx side, which are referred to as Rx impairments.

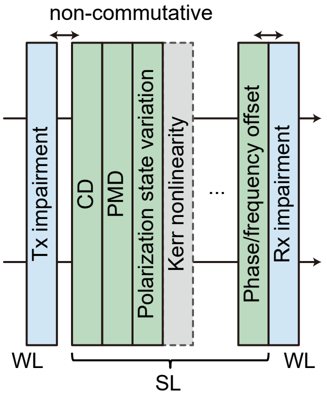


Figure 3.1. Impairments in coherent optical fiber communication system. CD: chromatic dispersion, PMD: polarization mode dispersion.

Except for nonlinearity in optical fiber propagation[1], these impairments described above are all linear processes. Tx and Rx impairments affect IQ components differently, and thus they are modeled as WL processes. In contrast, CD and PMD with polarization state variation in optical fiber propagation are SL (MIMO) processes. The phase/frequency offset is also an SL process. Since WL processes include SL processes as specific cases, the total impairment that includes all the relevant linear impairments shown in Fig. 3.1 is also modeled as a WL (MIMO) process. Therefore, a WL filter can compensate all the relevant

---

[1] Nonlinear distortion can occur in Tx and Rx devices, which are discussed in a later chapter.

linear impairments. However, this WL filter becomes redundantly large to have all the cross terms of IQ components of X/Y polarizations and cover the total temporal spread.

A conventional DSP for impairment compensation in optical fiber communication systems compensates for impairments in a block-wise manner [5]. This architecture enables efficient impairment compensation since the impairments have different characteristics. For example, CD in an optical fiber communication system can be regarded as static[2] and common for two polarizations, i.e., modeled with two single-input single-output processes for two polarizations. The temporal spread caused by CD can be huge in long-haul SMF transmission. Whereas, PMD with polarization state variation is a MIMO process of two polarizations and time-varying, which requires an adaptive filter to be compensated. The temporal spread caused by PMD is relatively small compared to that by CD.

Figure 3.2 shows a conventional DSP architecture for impairment compensation. It does not include Tx and Rx impairment compensation. CD compensation is performed first by static SL filters for couplex-valued received signals of two polarizations. Compensation for PMD and polarization state variation, which is referred to as polarization demultiplexing, is then performed by an adaptive MIMO SL filter. Carrier recovery is performed to compensate for the phase/frequency offset after or at the inner of polarization demultiplexing. Polarization demultiplexing and carrier recovery is usually based on an error of output symbols from the desired symbols in the IQ constellation. The appropriate error is not accessible when large CD is accumulated on the signal. Thus, polarization demultiplexing and carrier recovery are placed after static CD compensation.

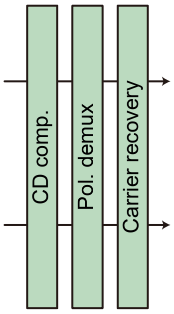


Figure 3.2. Conventional DSP architecture for impairment compensation.

A conventional DSP architecture shown in Fig. 3.2 consists of SL filters only. Therefore, it cannot compensate for Tx or Rx impairments which are WL processes. A WL filter or its equivalent real-valued MIMO filter of IQ components is required to compensate for Tx or Rx impairments [42]. Moreover, since Tx and Rx impairments depend on the device imperfection, the response are usually unknown and can be varied, for example, by the environment temper-

---

[2] If an optical path is fixed.

ature. The situation becomes further complicated since SL impairments occur in optical fiber propagation between the Tx and Rx, as shown in Fig. 3.1.

For these reasons, various approaches have been investigated to compensate for Tx or Rx impairments in optical fiber transmission. In order to compensate for Tx impairments, a real-valued MIMO filter for IQ components is placed after carrier recovery for each polarization [54]. The filter coefficient is adaptively controlled on the basis of the error of the filter output. In order to compensate for Rx impairments, CD compensation is performed independently on I and Q components of two polarizations to avoid IQ mixing, and a complex-valued adaptive 4×2 MIMO filter with the inputs of IQ components after independent CD compensation is used [9]. An adaptive WL filter is also used to compensate for Rx impairments together with CD [42]. In these approaches, either Tx impairments or Rx impairments are compensated adaptively. For compensation of both Tx and Rx impairments individually, ellipse correction and a k-means clustering with a blind phase search [55] are used for compensation of Rx impairments and Tx impairments, respectively, based on the features of signals with IQ impairments in the IQ constellation plane [56]. In [57], IQ phase deviation and IQ imbalance in the Rx are compensated by Gram-Schmidt orthogonal procedure [8], IQ skew in the Rx is compensated by a complex-valued adaptive 4×2 filter with independent CD compensation for IQ components. Tx impairments are compensated by two cascaded real-valued adaptive 4×4 MIMO filters for IQ components of two polarizations. In a similar approach, two adaptive 4×2 WL filters are used for compensation of Tx and Rx impairments [58].

In these previous approaches, Tx impairment compensation and Rx impairment compensation are performed in independent blocks[3] in a block-wise manner. When Tx and Rx impairments simultaneously occur in an optical fiber communication system, which is a normal situation, it is inevitable for at least one compensation block to work under the remaining impairments. In addition, CD compensation is required first before an adaptive filter to compensate for Rx impairments[4] because adaptive control of filter coefficients is based on the features of a transmitted signal[5] and it does not work under signal distortion due to large CD accumulation. Independent CD compensation for the IQ components of the received signal is adopted to avoid IQ mixing caused by CD compensation. It doubles the computational complexity of CD compensation. Temporal spread due to CD can be very large in long-haul SMF transmission compared to other impairments, and thus CD compensation requires a filter with a large temporal spread.

Another approach to compensate for Tx and Rx impairments simultaneously, an adaptive complex-valued 8×2 filter with independent CD compensation on IQ components and augmented phase/frequency offset compensation is used [59, 60]. This approach can compensate for Tx and Rx impairments adaptively and precisely, though it requires an enlarged adaptive MIMO filter resulting in higher computational complexity.

---

[3] Filter coefficients of Tx and Rx impairment compensation are controlled within individual blocks.

[4] Otherwise, CD is compensated together with Rx impairments by an adaptive WL filter.

[5] That is, a known transmitted symbol set like QAM or a constant amplitude of transmitted symbols.

3.1.1 Mutual commutativity of multiple impairments

Tx and Rx impairment compensation requires complicated filter structures described above because of the mutual non-commutativity of impairments and their compensation. As we have reviewed, primal linear impairments in optical fiber communication systems can be modeled by the multiplication of a transfer matrix in the frequency domain[6]. Impairments in the optical domain are MIMO processes since an SMF supports two polarization modes. Those in the electrical domain are also intrinsically MIMO processes of IQ components. Therefore, multiple impairments are not mutually commutative in general, as the multiplication of two matrices is not mutually commutative in general, which indicates that we should consider the order of impairments carefully.

Some impairment is commutative with other impairments. For example, the frequency response of CD, which is SL, represented by Eq. (2.43) in the complex-valued phasor representation is diagonal, and thus it is commutative with PMD whose frequency response is represented by Eq. (2.47). This fact justifies the conventional block-wise DSP for impairment compensation in which accumulated CD is compensated first and then PMD and the polarization state variation are compensated [5,61,62]. In contrast, if we focus on one polarization and consider the frequency response of CD in the IQ representation according to Eq. (2.11), it becomes as

$$H^{\mathbb{R}}_{\mathrm{CD}}(\omega) = \frac{1}{2}T^{\dagger}\begin{pmatrix} H_{\mathrm{CD}}(\omega) & 0 \\ 0 & H^{*}_{\mathrm{CD}}(\omega) \end{pmatrix} T = \begin{pmatrix} \mathrm{Re}[H_{\mathrm{CD}}(\omega)] & -\mathrm{Im}[H_{\mathrm{CD}}(\omega)] \\ \mathrm{Im}[H_{\mathrm{CD}}(\omega)] & \mathrm{Re}[H_{\mathrm{CD}}(\omega)] \end{pmatrix}, \tag{3.1}$$

where we use $H_{*\mathrm{CD}}(\omega) = 0$ and $H_{\mathrm{CD}}(-\omega) = H_{\mathrm{CD}}(\omega)$. This frequency response of CD in the IQ representation is no longer diagonal. The input IQ components are complicatedly mixed after CD accumulation, which is referred to as IQ mixing [9]. In consequence, CD cannot generally be commutative with another WL process such as IQ skew. For example, when IQ impairments that occur in a Tx and Rx, which are WL, are compensated in an Rx-side DSP in an optical fiber communication system with large accumulated CD, the compensation order is essential. The conventional DSP configuration shown in Fig. 3.2 is justified because CD is commutative with PMD and polarization state variation[7]. The mutual non-commutativity of impairments and compensation of them provides a critical role in this thesis.

3.2 Adaptive multi-layer strictly linear and widely linear filters

Since there are multiple impairments that are not mutually commutative in an optical fiber communication system, impairment compensation should be executed in the reverse order in which the impairments occur. Considering the mutual commutativity of impairments, an ideal DSP configuration to compensate for impairments in the reverse order in which they occur when any of the two are not mutually commutative is shown in Fig. 3.3. In this configuration,

---

[6] This is a consequence of the linear time-invariant system.

[7] When nonlinearity in optical fiber propagation cannot be ignored, this is no longer valid. A clear example is found in fiber nonlinearity compensation by a split-step method [7].

Rx impairment compensation is placed before CD compensation, which makes it difficult to adaptively optimize the coefficients within the Rx impairment compensation block locally since CD is accumulated on the signal.

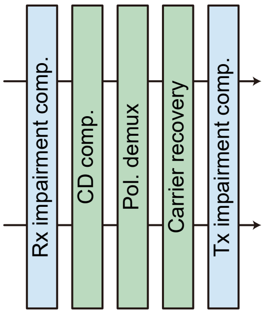


Figure 3.3. Ideal DSP configuration to compensate for impairments in the reverse order in which they occur when any of two are not mutually commutative.

In order to resolve this problem, gradient descent and gradient calculation with back propagation from the output through multiple cascaded blocks, i.e., layers, are adopted in this study. Gradient calculation with back propagation is widely used in machine learning literature with neural networks [63], while it can be applied to a computational graph in which elements are connected with differentiable input-output relations [37]. From this point of view, the output of an SL or WL filter is differentiable in terms of its input and filter coefficients. By applying gradient calculation with back propagation to multi-layered filters like Fig. 3.3, we can optimize the filter coefficients in all the layers to minimize the loss function consisting of the final output, in which all the relevant impairments should be compensated, without relying on local optimization of each middle layer. Precise compensation of multiple impairments is achieved by non-local optimization. Moreover, filters in multiple layers can be designed to compensate for the specific impairment(s) by considering the characteristics of the impairment(s) such as SL or WL, and no cross terms or MIMO[8], which enables efficient Tx and Rx impairment compensation in terms of computational complexity while using conventional SL CD compensation and polarization demultiplexing.

Figure 3.4 shows the multi-layer SL and WL MIMO filters corresponding to impairment compensation shown in Fig. 3.3 to compensate for all the linear impairments including Tx and Rx impairments. The multi-layer filters are composed of five layers for Rx impairment compensation, CD compensation, polarization demultiplexing, carrier recovery, and Tx impairment compensation, in this order. We regard all the filters as half-symbol spaced FIR filters for simplicity.

The first layer consists of two 2×1 WL filters for two polarizations[9] to com-

---

[8] Arrangement and design of filters can be regarded as an example of taking domain-specific knowledge of optical fiber communication systems into account for impairment compensation.

[9] We can regard them as a WL MIMO filter with no polarization cross-terms.

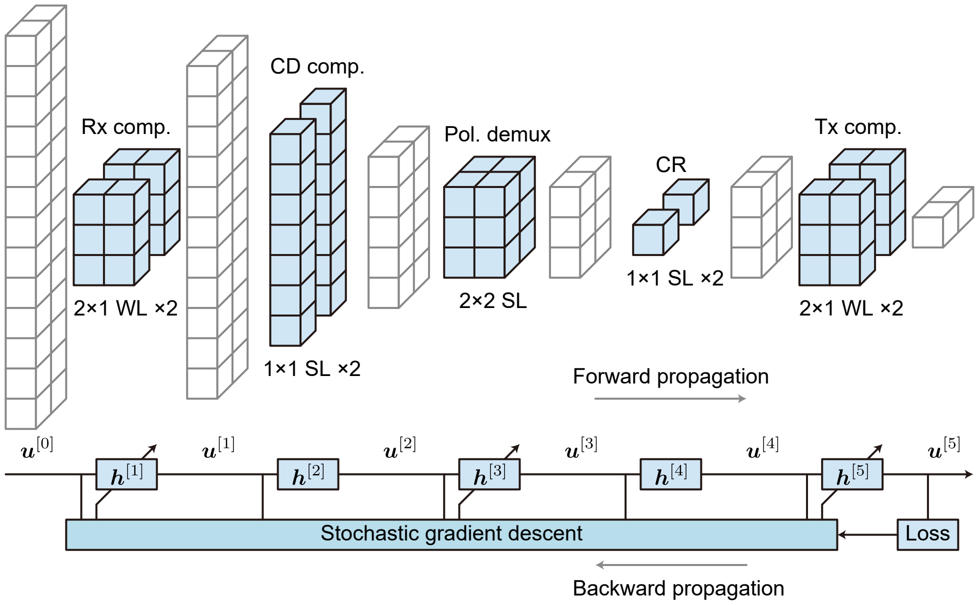


Figure 3.4. Adaptive multi-layer SL&WL filter architecture for compensation of Tx and Rx impairments. The filter coefficients in all layers are controlled by stochastic gradient descent with back propagation.

pensate for IQ impairments in Rx whose coefficients are controlled adaptively. The second layer consists of two 1×1 SL filters for CD compensation whose coefficients are static. The third layer consists of a 2×2 SL MIMO filter for polarization demultiplexing whose coefficients are adaptively controlled fast enough to track the polarization state variation in fiber propagation. The fourth layer consists of two 1×1 1-tap SL filters to compensate for the phase and frequency offset. The fifth layer consists of two adaptive 2×1 WL filters for compensation of IQ impairments in Tx. In this multi-layer FIR filter architecture, the input samples of longer duration are related to the final outputs in one symbol time slot, going back over the filter layer.

The filter coefficients of the second layer $\boldsymbol{h}_i^{[2]}$ ($i = 1, 2$, or X, Y for two polarizations) for CD compensation are set on the basis of the physical model of CD as[10]

$$\boldsymbol{h}_i^{[2]} = \mathcal{F}^{-1}[H_{\mathrm{CDC}}[\omega]], \tag{3.2}$$

$$H_{\mathrm{CDC}}[\omega] = \exp\left(i\frac{cD_{\mathrm{CD}}z}{4\pi\nu^2}\omega^2\right), \tag{3.3}$$

where c, D_{CD}, z, and ν are the speed of light, the dispersion coefficient, the transmission distance, and the carrier frequency of the signal, respectively. The fourth layer $\boldsymbol{h}_i^{[4]}$ executes the phase rotation for carrier recovery as

$$h_i^{[4]}[k] = \exp(-i\theta_i[k]), \tag{3.4}$$

where k is the sample timing integer. The compensated phases $\theta_i[k]$ are determined by the digital phase-locked loop (PLL) based on the data-aided or decision-directed phase error of the output symbols of the last layer outside SGD[11].

The filter coefficients of the first, third, and fifth layers are adaptively controlled by SGD and gradient calculation with back propagation. The conventional SGD approach for one adaptive filter is extended here to multi-layer SL&WL filters. In the following, we describe the input-output relation, or forward propagation, of the SL and WL (MIMO) filters in each layer. Then we derive back propagation for the gradient calculation.

3.2.1 Forward propagation

We consider the filter of the l-th layer. In the multi-layer FIR filter architecture, the samples of longer duration are related to the outputs of the last layer in one time slot. The related output and input signals of the l-th layer to obtain the last layer's output samples at timing integer k are given by

$$\boldsymbol{u}_i^{[l]}[k] = (u_i^{[l]}[k;k], u_i^{[l]}[k-1;k], \ldots, u_i^{[l]}[k-M_l+1;k])^{\mathrm{T}}, \tag{3.5}$$

and

$$\boldsymbol{u}_i^{[l-1]}[k] = (u_i^{[l-1]}[k;k], u_i^{[l-1]}[k-1;k], \ldots, u_i^{[l-1]}[k-M_{l-1}+1;k])^{\mathrm{T}}, \tag{3.6}$$

---

[10] In the DSP, the discrete-time and frequency are treated. We use the notation $[x]$ for a discrete variable x.

[11] It is reported that carrier recovery can be included in control with SGD [64].

where M_l and M_{l-1} are the length of the output and input signal vectors. The input of the l-th layer corresponds to the output of the $(l-1)$-th layer. $u_i^{[l]}[k-m;k]$ denotes the m-th component of the signal vector obtained by the filter coefficients at timing integer k. The length of the l-th layer's output signals M_l, the length of its input signals M_{l-1}, and the filter tap length $M^{[l]}$ satisfy the following relation:

$$M^{[l]} = M_{l-1} - M_l + 1, \tag{3.7}$$

because of convolution. By this relation, the input signal length progressively increases toward the beginning layers.

Strictly linear MIMO FIR filter

When the filter of the l-th layer is an SL MIMO filter, its output samples are described as

$$u_i^{[l]}[k] = \sum_{j=1}^{K} \boldsymbol{h}_{ij}^{[l]\mathrm{T}}[k]\bar{\boldsymbol{u}}_j^{[l-1]}[k], \tag{3.8}$$

where $\bar{\boldsymbol{u}}_j^{[l-1]}[k] = (u_j^{[l-1]}[k], u_j^{[l-1]}[k-1], \ldots, u_j^{[l-1]}[k-M^{[l]}+1])^{\mathrm{T}}$. $K=2$ corresponding to two polarizations. This includes the case of two 1×1 filters if removing the summation over j. The filter coefficients at the timing integer k are

$$\boldsymbol{h}_{ij}^{[l]}[k] = (h_{ij}^{[l]}[0;k], h_{ij}^{[l]}[1;k], \ldots, h_{ij}^{[l]}[M^{[l]}-1;k])^{\mathrm{T}}. \tag{3.9}$$

Using these descriptions, the output signal vectors of the l-th layer filter are described as

$$\boldsymbol{u}_i^{[l]}[k] = \sum_{j=1}^{K} H_{ij}^{[l]}[k]\boldsymbol{u}_j^{[l-1]}[k], \tag{3.10}$$

where

$$H_{ij}^{[l]}[k] = \begin{pmatrix} h_{ij}^{[l]}[0;k] & h_{ij}^{[l]}[1;k] & \cdots & h_{ij}^{[l]}[M^{[l]}-1;k] & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & & & & & & 0 \\ 0 & \cdots & 0 & h_{ij}^{[l]}[0;k] & h_{ij}^{[l]}[1;k] & \cdots & h_{ij}^{[l]}[M^{[l]}-1;k] \end{pmatrix} \tag{3.11}$$

is a matrix of the size $M_l \times M_{l-1}$. Equation (3.10) is rewritten as

$$\boldsymbol{u}_i^{[l]}[k] = \sum_{j=1}^{K} U_j^{[l-1]}[k]\boldsymbol{h}_{ij}^{[l]}[k], \tag{3.12}$$

where

$$U_j^{[l-1]}[k] = \begin{pmatrix} u_j^{[l-1]}[k;k] & u_j^{[l-1]}[k-1;k] & \cdots & u_j^{[l-1]}[k-M^{[l]}+1;k] \\ u_j^{[l-1]}[k-1;k] & u_j^{[l-1]}[k-2;k] & \cdots & u_j^{[l-1]}[k-M^{[l]};k] \\ \vdots & & & \vdots \\ u_j^{[l-1]}[k-M_l+1;k] & u_j^{[l-1]}[k-M_l;k] & \cdots & u_j^{[l-1]}[k-M_{l-1}+1;k] \end{pmatrix}, \tag{3.13}$$

is a matrix of the size $M_l \times M^{[l]}$.

When the l-th layer is for carrier recovery, the carrier phase and frequency offset are described via multiplication in the time domain rather than convolution. The filters for carrier recovery in the adaptive multi-layer filters are 1-tap SL filters. When the compensation values at timing integers ranging from $k - M_l + 1$ to k are

$$\boldsymbol{h}_i^{[l]}[k : k - M_l + 1] = (h_i^{[l]}[0; k], h_i^{[l]}[0; k-1], \ldots, h_i^{[l]}[0; k - M_l + 1])^{\mathrm{T}}, \quad (3.14)$$

the output signals of the l-th layer are given by

$$\boldsymbol{u}_i^{[l]}[k] = \boldsymbol{h}_i^{[l]}[k : k - M_l + 1] \circ \boldsymbol{u}_i^{[l-1]}[k]. \quad (3.15)$$

Consequently, the carrier phase and frequency offset or their compensation are not mutually commutative with other SL processes such as CD, strictly speaking, when the phase offset varies over time, which results in equalization-enhanced phase noise [65,66]. On the other hand, suppose that the phase offset can be assumed to be static and common to the two polarizations, which roughly holds for the carrier recovery filters in the adaptive multi-layer filters because they do not deal with a large temporal spread. In this case, the phase offset or its compensation is commutative with CD, but not commutative with WL processes such as Tx and Rx impairments.

Widely linear MIMO FIR filter

When the filter of the l-th layer is a WL MIMO filter, its output samples are described as

$$u_i^{[l]}[k] = \sum_{j=1}^{K} \boldsymbol{h}_{ij}^{[l]\mathrm{T}}[k] \bar{\boldsymbol{u}}_j^{[l-1]}[k] + \sum_{j=1}^{K} \boldsymbol{h}_{*ij}^{[l]\mathrm{T}}[k] \bar{\boldsymbol{u}}_j^{[l-1]*}[k]. \quad (3.16)$$

In the case of the WL MIMO filter, the filter coefficients are $\boldsymbol{h}_{ij}^{[l]}[k]$ and[12]

$$\boldsymbol{h}_{*ij}^{[l]}[k] = (h_{*ij}^{[l]}[0; k], h_{*ij}^{[l]}[1; k], \ldots, h_{*ij}^{[l]}[M^{[l]} - 1; k])^{\mathrm{T}}. \quad (3.17)$$

The output signal vectors are described as

$$\boldsymbol{u}_i^{[l]}[k] = \sum_{j=1}^{K} H_{ij}^{[l]}[k] \boldsymbol{u}_j^{[l-1]}[k] + \sum_{j=1}^{K} H_{*ij}^{[l]}[k] \boldsymbol{u}_j^{[l-1]*}[k], \quad (3.18)$$

where

$$H_{*ij}^{[l]}[k] = \begin{pmatrix} h_{*ij}^{[l]}[0;k] & h_{*ij}^{[l]}[1;k] & \cdots & h_{*ij}^{[l]}[M^{[l]}-1;k] & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & & & & & & 0 \\ 0 & \cdots & 0 & h_{*ij}^{[l]}[0;k] & h_{*ij}^{[l]}[1;k] & \cdots & h_{*ij}^{[l]}[M^{[l]}-1;k] \end{pmatrix}. \quad (3.19)$$

---

[12] If we denote $\boldsymbol{h}_{*ij}$ as $\boldsymbol{h}_{ij*}$, the subscript can be a bit constructive. Namely, $\boldsymbol{h}_{ij*}$ corresponds to the response from the conjugate of the j-th input signal ($j*$) to the i-th output signal.

It can be rewritten as

$$\boldsymbol{u}_i^{[l]}[k] = \sum_{j=1}^{K} U_j^{[l-1]}[k]\boldsymbol{h}_{ij}^{[l]}[k] + \sum_{j=1}^{K} U_j^{[l-1]*}[k]\boldsymbol{h}_{*ij}^{[l]}[k]. \tag{3.20}$$

By applying the above forward propagation from the first layer to the last (L-th) layer, we obtain the outputs $u_i^{[L]}[k]$ of the adaptive filters at timing integer k. The length of the last layer's outputs is $M_L = 1$.

3.2.2 Loss function

We consider here the case when the timing integer k denotes the symbol timing. The loss function ϕ to be minimized is constructed by the outputs $u_i^{[L]}[k]$ of the last L-th layer of the adaptive multi-layer SL&WL filters. The coefficient $\xi \in \mathbb{C}$ in any filter is controlled with SGD. Since we are dealing with complex-valued signals and filter coefficients, we utilize Wirtinger derivatives. Gradient descent becomes

$$\xi \to \xi - 2\alpha \frac{\partial \phi}{\partial \xi^*}, \tag{3.21}$$

where α is the step size. Since the loss function to be minimized should be real-valued,

$$\frac{\partial \phi}{\partial \xi^*} = \left(\frac{\partial \phi}{\partial \xi}\right)^*, \tag{3.22}$$

which also holds for the gradient in terms of the signal $\boldsymbol{u}_i^{[l]}[k]$.

There are several criteria for the loss function for adaptive filters. The CMA exploits the feature of the constant amplitude of the transmitted symbols[13] [5, 68]. In this case, the instantaneous loss function[14] is the squared magnitude of the error of the outputs from the constant amplitude r, as

$$\phi[k] = \sum_{i=1}^{K} (r^2 - |u_i^{[L]}[k]|^2)^2. \tag{3.23}$$

The gradients in terms of the last outputs $u_i^{[L]}[k]$ are calculated as

$$\frac{\partial \phi}{\partial u_i^{[L]}[k]} = -2e_i[k]u_i^{[L]*}[k], \tag{3.24}$$

where $e_i[k] = r^2 - |u_i^{[L]}[k]|^2$.

Data-aided or decision-directed least mean square (LMS) algorithm uses the feature that the transmitted symbols are the elements of the set determined by

---

[13] It is strictly true for some modulation formats, i.e., QPSK. CMA is usually used for them. Several extended versions exist for non-constant modulation formats, such as the multi-modulus algorithm for QAM [67].

[14] A mean squared error averaged over infinite or very long symbol periods should be a target to truly minimize, though it is hard to access for DSP. SGD minimizes the instantaneous error at one symbol period instead of the true mean error. This corresponds to mini-batch learning in a neural network literature with one batch size.

the modulation format itself [5]. The instantaneous loss function is the squared error of the outputs from the symbol candidates d as

$$\phi[k] = \sum_{i=1}^{K} |d_i[k] - u_i^{[L]}[k]|^2, \tag{3.25}$$

where d_i are given by the pre-known pilot patterns in the data-aided[15] LMS algorithm, or by results of decision operation to the outputs $u_i^{[L]}[k]$. The gradients in terms of the last outputs are

$$\frac{\partial \phi}{\partial u_i^{[L]}[k]} = -e_i[k]^*, \tag{3.26}$$

where $e_i[k] = d_i[k] - u_i^{[L]}[k]$.

3.2.3 Back propagation

Given the gradients in terms of the outputs of the filter of the l-th layer, we can use back propagation to calculate the gradients in terms of its inputs, or the outputs of the $(l-1)$-th layer. The gradients in terms of its filter coefficients can also be calculated with back propagation. We derive back propagation for the SL and WL MIMO FIR filters.

Strictly linear MIMO FIR filter

After calculating differentials and arranging them, the gradients of the loss function ϕ in terms of the inputs and filter coefficients of an SL MIMO filter are

$$\frac{\partial \phi}{\partial \boldsymbol{u}_j^{[l-1]}[k]} = \sum_{i=1}^{K} H_{ij}^{[l]\mathrm{T}}[k] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]}[k]}, \tag{3.27}$$

$$\frac{\partial \phi}{\partial \boldsymbol{h}_{ij}^{[l]}[k]} = U_j^{[l-1]\mathrm{T}}[k] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]}[k]}. \tag{3.28}$$

As for the case when the l-th layer consists of 1-tap SL filters for carrier recovery, the gradients in terms of the inputs are

$$\frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l-1]}[k]} = \boldsymbol{h}_i^{[l]}[k : k - M_l + 1] \circ \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]}[k]}. \tag{3.29}$$

Here, note that the filter coefficients for carrier recovery are not controlled by gradient descent in the adaptive multi-layer filters in this study.

---

[15] An adaptive control of the filter coefficients based on pre-known symbols for the Rx is referred to as data-aided. It requires the insertion of pilot symbols to the transmitted symbols in the Tx, which cannot carry the data to be communicated. Antonym is blind, which works without knowing what symbols are transmitted. CMA and decision-directed LMS are categorized as blind.

Widely linear MIMO FIR filter

Similarly, the gradients of the loss function ϕ in terms of the inputs and filter coefficients of a WL MIMO filter are

$$\frac{\partial \phi}{\partial \boldsymbol{u}_j^{[l-1]}[k]} = \sum_{i=1}^{K} \left(H_{ij}^{[l]\mathrm{T}}[k] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]}[k]} + H_{*ij}^{[l]\dagger}[k] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]*}[k]} \right), \tag{3.30}$$

$$\frac{\partial \phi}{\partial \boldsymbol{h}_{ij}^{[l]}[k]} = U_j^{[l-1]\mathrm{T}}[k] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]}[k]}, \tag{3.31}$$

$$\frac{\partial \phi}{\partial \boldsymbol{h}_{*ij}^{[l]}[k]} = U_j^{[l-1]\dagger}[k] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]}[k]}. \tag{3.32}$$

Now, we have an adaptive filter coefficient update algorithm for all the filters. The gradients of the loss function in terms of the filter coefficients back to the first layer can be calculated from the last l-the layer with these back propagation equations successively.

3.2.4 Approximate calculation

We give an approximated calculation for adaptive multi-layer filters. Because the adaptive multi-layer filters deal with time-series signals and a coefficient update is usually small in one update of SGD, a portion of the intermediate outputs of layers other than the last layer can be approximately reused after a coefficient update. In contradistinction to this approximate calculation, we refer to the calculation of forward and back propagation described so far as full calculation. This approximation reduces the computational complexity for forward propagation of the layers to that of a conventional FIR filter.

Let us consider forward propagation of the multi-layer filters at timing integer $k+1$ after a coefficient update at timing integer k. From the full calculation above, the output signals of the l-th layer at timing integer $k+1$ are given by

$$\boldsymbol{u}_i^{[l]}[k+1] = (u_i^{[l]}[k+1; k+1], u_i^{[l]}[k; k+1], \ldots, u_i^{[l]}[k - M_l + 2; k+1])^{\mathrm{T}}, \tag{3.33}$$

except when the l-th layer is for carrier recovery. Here, $\boldsymbol{u}_i^{[l]}[k+1]$ does not share any elements with $\boldsymbol{u}_i^{[l]}[k]$. However, because the input signals are time series and the step sizes for coefficient updates are usually small for stochastic gradient descent, we can assume that

$$u_i^{[l]}[k-m; k+1] \sim u_i^{[l]}[k-m; k-m], \tag{3.34}$$

for an integer m. By this approximation, the output signals of the l-th layer become

$$\boldsymbol{u}_i^{[l]}[k+1] = (u_i^{[l]}[k+1; k+1], u_i^{[l]}[k; k], \ldots, u_i^{[l]}[k - M_l + 2; k - M_l + 2])^{\mathrm{T}}, \tag{3.35}$$

in which case they shares elements other than the first element with $\boldsymbol{u}_i^{[l]}[k]$. As a result, we only need to calculate $u_i^{[l]}[k+1; k+1]$ in forward propagation of the l-th layer at timing integer $k+1$.

Forward propagation

We consider forward and back propagation with the approximation of Eq. (3.35) in the case in which the l-th layer consists of a WL MIMO filter. Other type of filters correspond to a limited case of the WL MIMO filter. We separately examine the case for carrier recovery later. Under the approximation, the forward propagation in the case when the l-th layer consists of a WL MIMO filter is given by

$$\boldsymbol{u}_i^{[l]}[k] = \sum_{j=1}^{K} H_{ij}^{[l]}[k:k-M_l+1]\boldsymbol{u}_j^{[l-1]}[k] + \sum_{j=1}^{K} H_{*ij}^{[l]}[k:k-M_l+1]\boldsymbol{u}_j^{[l-1]*}[k], \tag{3.36}$$

where

$$H_{ij}^{[l]}[k:k-M_l+1] = \begin{pmatrix} h_{ij}^{[l]}[0;k] & \cdots & h_{ij}^{[l]}[M^{[l]}-1;k] & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ 0 & \cdots & 0 & h_{ij}^{[l]}[0;k-M_l+1] & \cdots & h_{ij}^{[l]}[M^{[l]}-1;k-M_l+1] \end{pmatrix}, \tag{3.37}$$

$$H_{*ij}^{[l]}[k:k-M_l+1] = \begin{pmatrix} h_{*ij}^{[l]}[0;k] & \cdots & h_{*ij}^{[l]}[M^{[l]}-1;k] & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ 0 & \cdots & 0 & h_{*ij}^{[l]}[0;k-M_l+1] & \cdots & h_{*ij}^{[l]}[M^{[l]}-1;k-M_l+1] \end{pmatrix}. \tag{3.38}$$

As described above, most elements can be reused from the results at timing integer $k-1$, which greatly mitigates the computational complexity for forward propagation of the l-th layer. The same discussion also holds when the l-th layer consists of SL filters or is without MIMO. Equation (3.36) also handles the case when the l-th layer consists of filters for carrier recovery, because $H_{ij}^{[l]}[k:k-M_l+1] = \mathrm{diag}(\boldsymbol{h}_i^{[l]}[k:k-M_l+1])$ in this case.

Back propagation

Next the back propagation in the case when the l-th layer consists of a WL MIMO filter is calculated as follows. The gradients of the loss in terms of the inputs are given by

$$\frac{\partial \phi}{\partial \boldsymbol{u}_j^{[l-1]}[k]} = \sum_{i=1}^{K} \left(H_{ij}^{[l]\mathrm{T}}[k:k-M_l+1] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]}[k]} + H_{*ij}^{[l]\dagger}[k:k-M_l+1] \frac{\partial \phi}{\partial \boldsymbol{u}_i^{[l]*}[k]} \right). \tag{3.39}$$

For the gradients in terms of the filter coefficients, although we can calculate $\partial\phi/\partial\boldsymbol{h}_{ij}^{[l]*}[k]$ and $\partial\phi/\partial\boldsymbol{h}_{*ij}^{[l]*}[k]$ for the coefficients at timing integer k, these expressions do not work for gradient descent. They include $\partial\phi/\partial u_i^{[l]*}[k]$, but this term can be near zero because the magnitude of the first coefficient $h_{ij}^{[l]}[0;k]$ or $h_{*ij}^{[l]}[0;k]$ for a conventional causal FIR filter can be very small in Eqs. (3.37) and (3.38). Accordingly, instead of obtaining the gradients in terms of the coefficients at timing integer k, we modify the coefficient update by using the gradients in terms of the filter coefficients that are related in calculating all the output signals of the l-th layer, as follows:

$$\boldsymbol{h}_{ij}^{[l]}[k] \to \boldsymbol{h}_{ij}^{[l]}[k] - 2\alpha \sum_{m=0}^{M_l-1} \frac{\partial\phi}{\partial\boldsymbol{h}_{ij}^{[l]*}[k-m]}, \tag{3.40}$$

$$\boldsymbol{h}_{*ij}^{[l]}[k] \to \boldsymbol{h}_{*ij}^{[l]}[k] - 2\alpha \sum_{m=0}^{M_l-1} \frac{\partial\phi}{\partial\boldsymbol{h}_{*ij}^{[l]*}[k-m]}. \tag{3.41}$$

In this case, the summations of the gradients are written as

$$\sum_{m=0}^{M_l-1} \frac{\partial\phi}{\partial\boldsymbol{h}_{ij}^{[l]*}[k-m]} = U_j^{[l-1]\dagger}[k:k-M_{l-1}+1]\frac{\partial\phi}{\partial\boldsymbol{u}_i^{[l]*}[k]}, \tag{3.42}$$

$$\sum_{m=0}^{M_l-1} \frac{\partial\phi}{\partial\boldsymbol{h}_{*ij}^{[l]*}[k-m]} = U_j^{[l-1]\mathrm{T}}[k:k-M_{l-1}+1]\frac{\partial\phi}{\partial\boldsymbol{u}_i^{[l]*}[k]}, \tag{3.43}$$

where

$$\begin{aligned}&U_j^{[l-1]}[k:k-M_{l-1}+1] = \\ &\begin{pmatrix} u_j^{[l-1]}[k;k] & \cdots & u_j^{[l-1]}[k-M^{[l]}+1;k-M^{[l]}+1] \\ u_j^{[l-1]}[k-1;k-1] & \cdots & u_j^{[l-1]}[k-M^{[l]};k-M^{[l]}] \\ \vdots & & \vdots \\ u_j^{[l-1]}[k-M_l+1;k-M_l+1] & \cdots & u_j^{[l-1]}[k-M_{l-1}+1;k-M_{l-1}+1] \end{pmatrix}.\end{aligned} \tag{3.44}$$

In this way, we can obtain the back propagation for gradient descent to control the filter coefficients of the multi-layer filters in the case of the approximation.

3.2.5 Computational complexity

The indicator of the computational complexity of the DSP for communication systems is usually the required number of (complex-valued) multiplications. We first estimated the computational complexities of forward propagation and back propagation for the full and approximated calculations in the case when the l-th layer consists of a WL MIMO filter. The number of complex multiplications required in the l-th layer for the last layer's outputs at one timing integer was estimated.

The results are listed in Table 3.1. The full calculation requires $2K^2M^{[l]}M_l$ complex multiplications for forward propagation, whereas the approximated calculation requires $2K^2M^{[l]}$ complex multiplications, which corresponds to the convolution for the $M^{[l]}$-tap WL MIMO filter itself. As for back propagation, the full and approximated calculations require the same number of complex multiplications: $2K^2M^{[l]}M_l$ for the inputs and $2K^2M^{[l]}M_l$ for the coefficients, giving a total of $4K^2M^{[l]}M_l$. Forward propagation of the multi-layer filters is necessary for every symbol integer, whereas back propagation can be performed intermittently if the temporal variation of the impairments in question is slow. Therefore, the approximation's mitigation of the computational complexity for forward propagation of the multi-layer filters is significant.

Table 3.1: Number of the complex multiplications required in l-th layer for the last layer's outputs at one timing integer when the l-th layer consists of a WL MIMO filter.

	Forward	Back (signal)	Back (coefficient)
Full	$2K^2M^{[l]}M_l$	$2K^2M^{[l]}M_l$	$2K^2M^{[l]}M_l$
Approx.	$2K^2M^{[l]}$	$2K^2M^{[l]}M_l$	$2K^2M^{[l]}M_l$

The required complex multiplications in the l-th layer for the last layer's outputs at one timing integer with approximate calculation for respective filters are listed in Table 3.2. In the later evaluation in this chapter, we use the adaptive multi-layer filters consisting of five layers shown in 3.4 in the case with the filter lengths of $M^{[1]} = 5$, $M^{[2]} = 61$, $M^{[3]} = 21$, $M^{[4]} = 1$, $M^{[5]} = 5$. In this case, $M_0 = 89$, $M_1 = 85$, $M_2 = 25$, $M_3 = 5$, $M_4 = 5$, $M_5 = 1$. Using the corresponding cases for layers in Table 3.2, forward propagation for the adaptive multi-layer SL&WL filter architecture requires 248 complex-valued multiplications. It requires 7340 complex-valued multiplications for back propagation straightforwardly, though it can be largely mitigated. Although Tx and Rx impairments vary in time, the speed is much slower than that of PMD and polarization state variation in an optical fiber since it depends on the environment temperature or aging of devices. Adaptive coefficient control is not necessarily required in an embedded real-time processor for all the layers. Only the gradient calculation for the third layer is required in real-time[16]. In this case, the gradient calculation for signals is required for the fourth and fifth layers, and the gradient calculation for coefficients is required for the third layers. It results in 450 complex-valued multiplications.

Table 3.2: Number of the complex multiplications required in l-th layer for the last layer's outputs at one timing integer with approximate calculation for respective filters.

	Forward	Back (signal)	Back (coefficient)
SL	$KM^{[l]}$	$KM^{[l]}M_l$	$KM^{[l]}M_l$
SL MIMO	K^2	$K^2M^{[l]}M_l$	$K^2M^{[l]}M_l$
WL	$2KM^{[l]}$	$2KM^{[l]}M_l$	$2KM^{[l]}M_l$
WL MIMO	$2K^2M^{[l]}$	$2K^2M^{[l]}M_l$	$2K^2M^{[l]}M_l$

---

[16] The coefficients of other layers can be determined by using computational resources outside of the embedded processor in the same way. Forward propagation is always necessary in real-time to obtain output signals.

For comparison, we examine the case of an adaptive 4×2 WL MIMO filter with the length corresponding to the total length of the adaptive multi-layer SL&WL filters described above[17]. When the adaptive 4×2 WL MIMO filter has a temporal length of 89 taps, it requires 712 complex-valued multiplications for forward propagation. It also requires 712 complex-valued multiplications for back propagation. Consequently, these results show that the multi-layer SL&WL filter architecture is more effective in compensating for all the relevant impairments than one large WL MIMO filter.

3.3 Evaluation in simulation

We evaluated the multi-layer SL&WL filters with SGD to compensate for all the relevant impairments by numerical simulation using a simple model. The reception of a 32-Gbaud PDM-QPSK signal with coherent detection was simulated. In Tx and Rx, IQ skew was imposed in the X polarization where the Q component was delayed. We assumed no polarization rotation, PMD, laser phase noise, or a frequency offset. After coherent reception, DSP was applied to the two-fold oversampled signals. We focused on IQ skew as Tx and Rx impairments.

We evaluated three types of DSP architecture shown in Fig. 3.5 for comparison. The first one shown in Fig. Fig. 3.5(a) corresponds to a conventional DSP for impairment compensation in optical fiber communications, as shown in Fig. 3.2, which is hereinafter referred to as 2×2 SL after the adaptive MIMO filter. CD compensation is performed on two polarization signals with SL filters. Polarization demultiplexing together with carrier recovery is performed by a 2×2 SL MIMO filter[18]. The second one shown in Fig. 3.5(b) corresponds to a conventional method with IQ skew compensation capability, which helps clarify the results. With this one, a 4×2 WL MIMO filter was used for the polarization demultiplexing block instead of a 2×2 SL MIMO filter (corresponding to the real-valued 4×4 filter for IQ components [9]), which is referred to as 4×2 WL. The third one shown in Fig. 3.5(c) was the multi-layer SL&WL filters with SGD, which we call Multi-layer SL&WL. CD compensation was performed by a 61-tap FIR filter enabling compensation of accumulated CD over 100-km SMF in all cases. The polarization demultiplexing block had the tap length of 21. In the case of Multi-layer SL&WL, Rx/Tx device compensation was performed by two 2×1 WL filters with five taps for two polarizations. The filter coefficients at the center of the main diagonal position of the MIMO FIR filters were initialized to one and the rest were zero. The loss function was decision-directed LMS.

---

[17] This configuration can compensate only for Rx impairments when a frequency offset exists. In order to compensate for Tx and Rx impairments simultaneously with an adaptive filter block, an 8×2 WL MIMO filter with augmented CD compensation and phase/frequency offset compensation is required.

[18] Since a phase/frequency offset causes the phase rotation of the signal, carrier recovery is placed after the adaptive MIMO filter to use the appropriate decision-directed error after it [69]. From the point of view of adaptive multi-layer filters, it can be regarded as a simple of SGD control through multiple blocks.

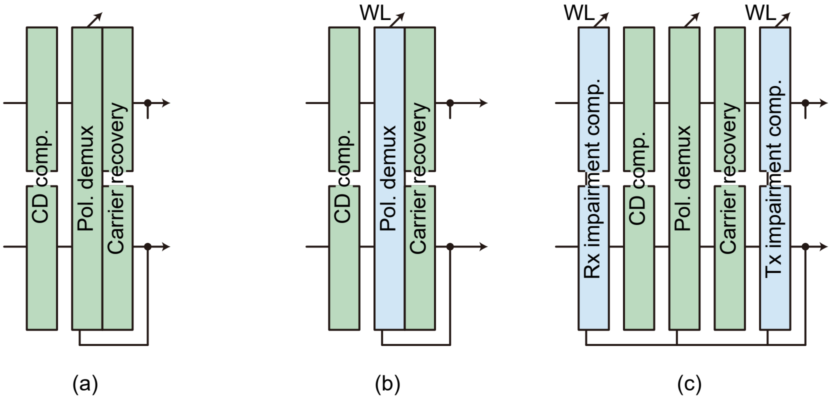


Figure 3.5. DSP architectures for evaluation: (a) conventional DSP for impairment compensation with static SL filters for CD compensation and an adaptive SL MIMO filter for polarization demultiplexing (2×2 SL), (b) DSP with static SL filters for CD compensation and an adaptive WL MIMO filter (4×2 WL), (c) adaptive multi-layer SL&WL filter architecture (Multi-layer SL&WL).

Back-to-back condition

The received constellations of the X and Y polarizations after DSP for impairment compensation under the back-to-back condition are shown in Fig. 3.6. The optical signal-to-noise ratio (OSNR) before the EDFA was set to 30 dB/0.1 nm. Figure 3.6(a) shows the simulation results under the condition with the X-IQ skew of 5 ps in Tx and the DSP architecture of 2×2 SL. Figure 3.6(b) shows that with the X-IQ skew of 5 ps in Rx and the DSP of 2×2 SL. Since the SL MIMO filter does not have the capability of IQ skew compensation, signal distortion occurred when there was IQ skew in either Tx or Rx. Figures 3.6(c) and (d) show the results with 4×2 WL with the X-IQ skew of 5 ps in Tx and Rx, respectively. In contrast to the case of 2×2 SL, IQ skew was compensated and good constellations were obtained. Figures 3.6(e) and (f) show the results with Multi-layer SL&WL with the X-IQ of 5 ps skew in Tx and Rx. IQ skew was fully compensated here as well.

100-km SMF transmission

The received constellations of the X and Y polarizations after impairment compensation DSP in 100-km SMF transmission, i.e., CD accumulation of the amount corresponding to 100 km of SMF between Tx and Rx, are shown in Fig. 3.7. The upper images (Figs. 3.7(a), (c), and (e)) are the results with the X-IQ skew of 5 ps in Tx, and the lower ones (Figs. 3.7(b), (d), and (f)) are those with the X-IQ skew of 5 ps in Rx. In the case of 2×2 SL (Figs 3.7(a) and (b)), the IQ skew was not compensated at all, the same as under the back-to-back condition. In the case of 4×2 WL (Figs. 3.7(c) and (d)), the IQ skew in Rx was not compensated, in contrast to the results under the back-to-back

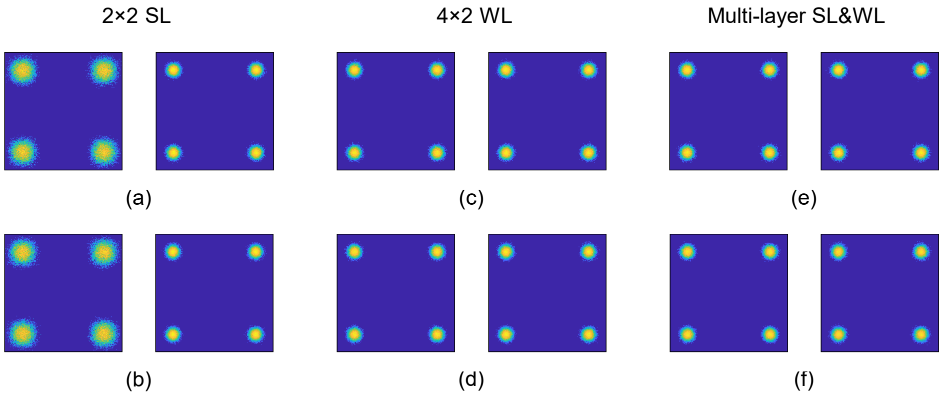


Figure 3.6. Simulation results for compensation of transmitter and receiver IQ skew under the back-to-back condition. Received constellations of PDM-QPSK by ((a) and (b)) 2×2 SL, by ((c) and (d)) 4×2 WL, and by ((e) and (f)) Multi-layer SL&WL. IQ skew of the X polarization signal was introduced at the transmitter side in (a), (c), and (e), and at the receiver side in (b), (d), and (f). The left side shows X polarization and the right side shows Y polarization.

condition shown in Fig. 3.7(d). This is because the 4×2 WL MIMO filter is not commutative to the CD compensation block (as can be seen in Fig. 3.1), which causes IQ mixing to occur in the CD compensation[19]. In contrast, in the case of Multi-layer SL&WL (shown in Figs 3.7(e) and (f)), good constellations were obtained even when the IQ skew in both Tx and Rx was imposed. This result demonstrates that the proposed multi-layer SL&WL filters with SGD can effectively compensate for IQ skew in both Tx and Rx under the accumulation of CD.

Compensation of simultaneous random Tx and Rx impairments

Whether the loss function of decision-directed LMS or data-aided LMS in terms of the adaptive filter coefficients is down-convex is not well known (or hard to guarantee to be down-convex), though polarization demultiplexing with an adaptive filter has been succeeded in optical fiber communications. Thus, to show that the multi-layer SL&WL filters with SGD works not only in a specific condition, we evaluated the performance in the case where IQ skew exists in all Tx/Rx and X/Y. The simulation model is the same as the previous one of 100-km transmission except that the received OSNR was set to 15 dB/0.1 nm. Each of Tx/Rx X/Y IQ skew was a random value from a zero-mean Gaussian distribution with a standard deviation equal to 5 ps. The averaged error vector magnitude (EVM) of the received signals over two polarizations was evaluated 100 times. Figure 3.8 shows the results of the histogram of the averaged EVM.

---

[19] If the CD compensation was applied to both I and Q independently to avoid IQ mixing through the CD compensation, the IQ skew in Rx can be compensated, while it requires additional large circuit resources.

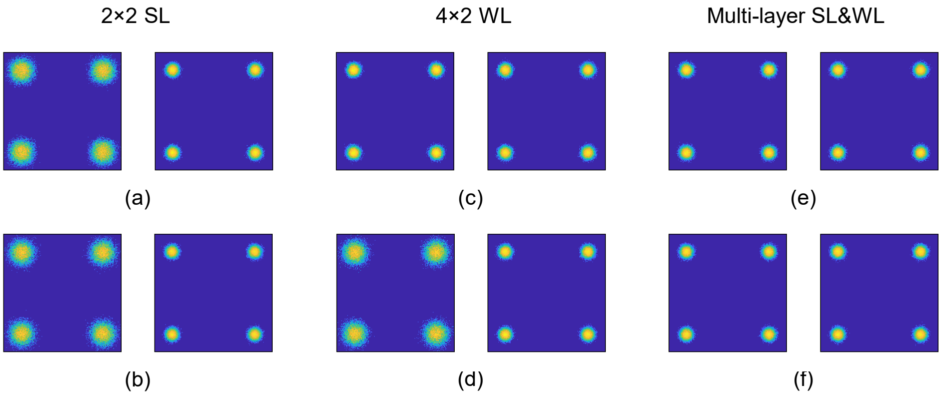


Figure 3.7. Simulation results for compensation of transmitter and receiver IQ skew after 100-km SMF transmission. Received constellations of PDM-QPSK by ((a) and (b)) 2×2 SL, by ((c) and (d)) 4×2 WL, and by ((e) and (f)) Multi-layer SL&WL. IQ skew of the X polarization signal was introduced at the transmitter side in (a), (c), and (e), and at the receiver side in (b), (d), and (f). The left side shows X polarization and the right side shows Y polarization.

Since 2×2 SL cannot compensate for IQ skew in both Tx and Rx and 4×2 WL cannot compensate for IQ skew in Rx in this condition, the averaged EVM with these two was spread and 4×2 WL provided slightly better results. In contrast, Multi-layer SL&WL provided good and almost constant averaged EVM, since it can compensate for IQ skew in both Tx and Rx. This result demonstrates that the Multi-layer SL&WL architecture works even in the complicated condition where IQ skew exists in all Tx/Rx and X/Y.

Convergence of loss

The convergence speed is one of the important characteristics of an adaptive filter. We compared the convergence speed in the cases of 2×2 SL, 4×2 WL, and Multi-layer SL&WL. The simulation model is the same as the one of 100-km SMF transmission of 32-Gbaud PDM-QPSK. The received OSNR was set to 15 dB/0.1 nm. Figure 3.9 shows the results of the time development of the loss function under the condition with the X-IQ skew of 5 ps in Rx. The loss was moving-averaged over 10 symbols. The convergence speed of Multi-layer SL&WL was similar to that of 2×2 SL and 4×2 WL in this case.

3.4 Evaluation in transmission experiment

Next, we experimentally evaluated the effectiveness of the proposed technique with 32-Gbaud PDM-64QAM. We focused on the capability of the IQ skew compensation in Tx and Rx, the same as in the simulation.

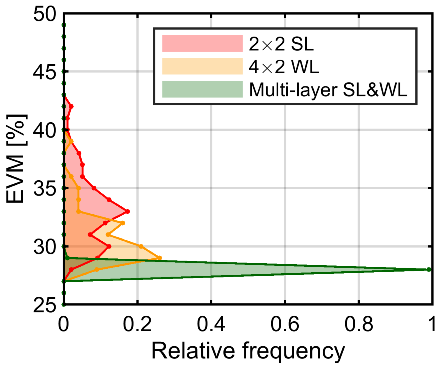


Figure 3.8. Simulation results of histogram of EVM with random IQ skew in all Tx/Rx and X/Y polarizations.

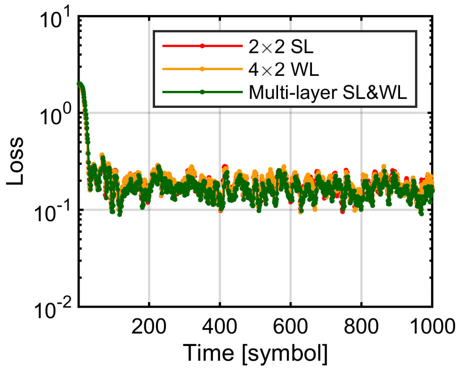


Figure 3.9. Simulation results of convergence of loss.

Experimental setup

A schematic diagram of the experimental setup is shown in Fig. 3.10. On the Tx side, a laser source at the frequency of 193.3 THz having a linewidth of about 100 kHz was modulated to 32-Gbaud PDM-64QAM by waveforms generated with a four-channel DAC at the sampling rate of 64 GS/s with a vertical resolution of eight bits. Forward error correction (FEC) of low-density parity-check code (LDPC) for DVB-S2 with a frame length of 64,800 and a code rate of 0.8 was used. Eight FEC frames were generated for each polarization by loading random bits to their payload and were then mapped to PDM-64QAM with gray mapping. In this experiment, a pilot sequence was inserted for each polarization to perform a pilot-based DSP in Rx [70, 71]. One pilot symbol of QPSK was inserted every 25 symbols. In addition, due to the restriction of the DAC used in the experiment, an overhead of QPSK symbols was also inserted to ensure the periodicity of the waveforms generated by DAC. QPSK symbols in the pilot and the overhead, which were about 7% in total, were set to the outer symbol points of 64QAM for simplicity of the decision processing in DSP, which yields 0.7 dB of the power penalty. The data generated in this way were upsampled to two-fold oversampling and the root raised cosine filter with a roll-off factor of 0.1 was performed. Frequency characteristics in the Tx devices were pre-compensated.

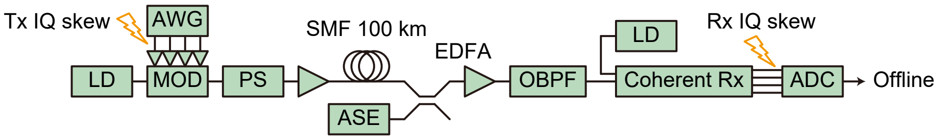


Figure 3.10. Experimental setup for compensation of transmitter and receiver IQ skew. PS: polarization scrambler, ASE: amplified spontaneous emission, OBPF: optical bandpass filter.

The modulated optical signal was transmitted to a 100-km SMF span after slow polarization scrambling at a rate of 10×2π rad/s. The span input optical power was set to 0 dBm. After transmission, amplified spontaneous emission was added to the signal to set an OSNR value. The optical signal was amplified by an EDFA and filtered by an optical bandpass filter having a 3-dB bandwidth of 50 GHz, and then received by a polarization diversity coherent receiver with a laser source used as an LO having a linewidth of about 100 kHz. The laser sources of the signal and the LO were free-running and had frequency differences fluctuated within about ±100 MHz. The four outputs of the coherent receiver were sampled with a digital oscilloscope at a sampling rate of 80 GS/s with a vertical resolution of eight bits. IQ skew in Tx and Rx was emulated digitally as a delay of the Q components of the X polarization.

DSP was performed offline. The received signals were normalized and resampled to two-fold oversampling. After matched root raised cosine filtering, detection of the pilot position and alignment was performed before the main DSP. This procedure was done in the following manner. First, CD compensation was applied to the received signals and the pilot position was detected according to the difference of the averaged powers between the 64QAM signal and the QPSK pilot. Then, polarization demultipexing was performed by CMA for the pilot. After frequency offset compensation based on the fourth power of the signal [72], timing alignment was done by using correlation to the known pilot sequence. The main DSP for impairment compensation, for which we evaluated the same three methods as in the previous simulation (2×2 SL, 4×2 WL, and Multi-layer SL&WL) was applied to the received signal after the matched filter with timing alignment before performing CD compensation again. In the experiment with PDM-64QAM, the filter coefficient update was carried out using the symbols of the QPSK pilot with data-aided LMS, with no updates by the 64QAM signal. The tap lengths of the filters were the same as in the simulation: 61 taps for CD compensation, 21 taps for polarization demultiplexing, and five taps for compensation of Tx/Rx device impairments. Carrier recovery was performed using both the signal and the pilot symbols with PLL based on a data-aided or decision-directed phase error and a second-order loop filter [73]. After the main DSP for impairment compensation and removal of the pilot and the overhead, the normalized generalized mutual information (NGMI) as a performance indicator was calculated [74], and the post FEC bit error rate (BER) was calculated with FEC decoding. The received waveforms were acquired three times under each condition and about 0.8 Mbits were evaluated for the post-FEC BER in each acquisition. The averaged NGMI was also evaluated. As the post-FEC BER largely fluctuated around the FEC cliff, we based our evaluation on its median.

Back-to-back condition

We first evaluated the capability of IQ skew compensation under the back-to-back condition. The received OSNR was set to 30 dB/0.1 nm. The results of the post-FEC BER while changing X-IQ skew from -10 ps to $+10$ ps in Tx and Rx are shown in Fig. 3.11(a) and (b), respectively. The error-free results are plotted at 10^{-5} for visibility. In the case of the IQ skew in Tx shown in Fig. 3.11(a), the proposed Multi-layer SL&WL could compensate for it, while 2×2 SL and 4×2 WL could not. In the case of the IQ skew in Rx shown in Fig. 3.11(b), 4×2 WL and Multi-layer SL&WL could compensate for it. In contrast to the simulation results (Fig. 3.6(c) and (d)), the IQ skew in Tx was not compensated by 4×2 WL, while IQ skew in Rx was. This is because a frequency offset existed in the experiment, rather than being set to zero as in the simulation, which renders the carrier recovery block and the Rx device compensation non-commutative, as shown in Fig. 3.1. Figures 3.12(a) and (b) show the corresponding evaluation results of the NGMI. The proposed Multi-layer SL&WL could compensate for IQ skew in both Tx and Rx. Figure 3.13 shows the received constellations after impairment compensation with Multi-layer SL&WL for several IQ skew values in Rx.

Figure 3.14 shows the dependence of the NGMI on IQ skew in Tx and Rx when they were added simultaneously. The 2×2 SL in Fig. 3.14(a) had no tolerance to IQ skew in either Tx or Rx. The 4×2 WL in Fig. 3.14(b) had tolerance to IQ skew in Tx but not in Rx. In contrast, Multi-layer SL&WL had tolerance to IQ skew in both Tx and Rx, and was able to deliver a stable performance.

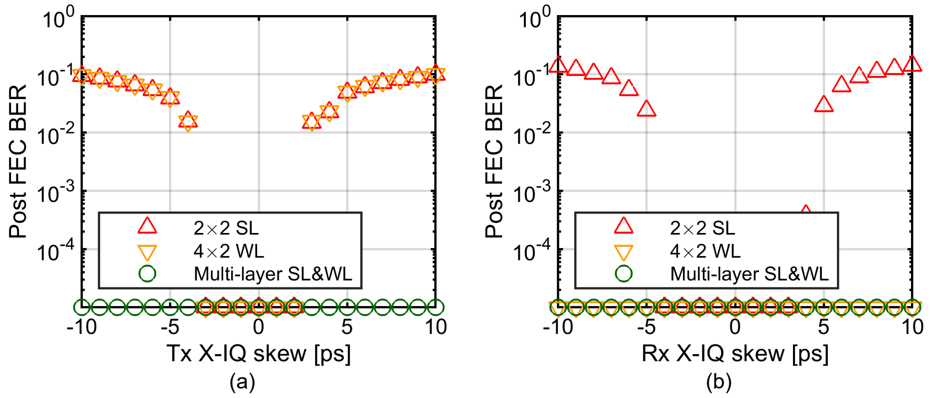


Figure 3.11. Experimental results for post-FEC BER under back-to-back condition with (a) Tx X-IQ skew and (b) Rx X-IQ skew.

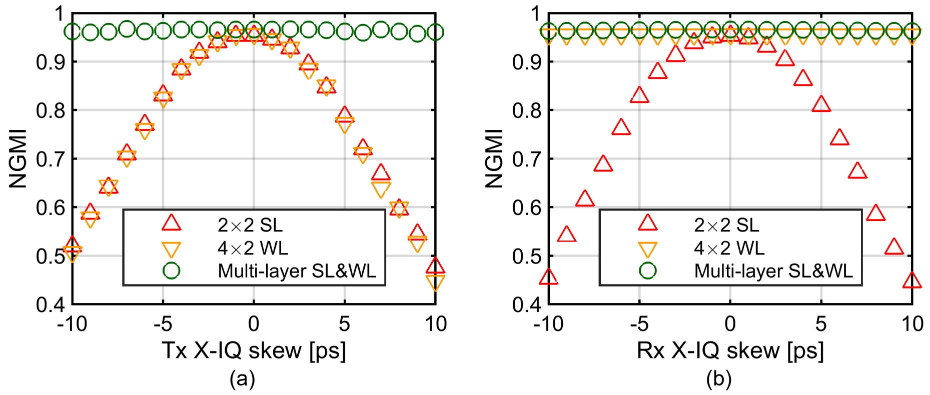


Figure 3.12. Experimental results for NGMI under back-to-back condition with (a) Tx X-IQ skew and (b) Rx X-IQ skew.

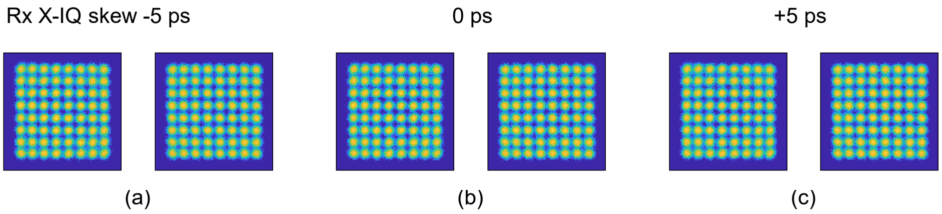


Figure 3.13. Received constellations after impairment compensation DSP with Multi-layer SL&WL under back-to-back condition with Rx X-IQ skew of (a) -5 ps, (b) 0 ps, and (c) +5 ps. The left side shows X polarization and the right side shows Y polarization.

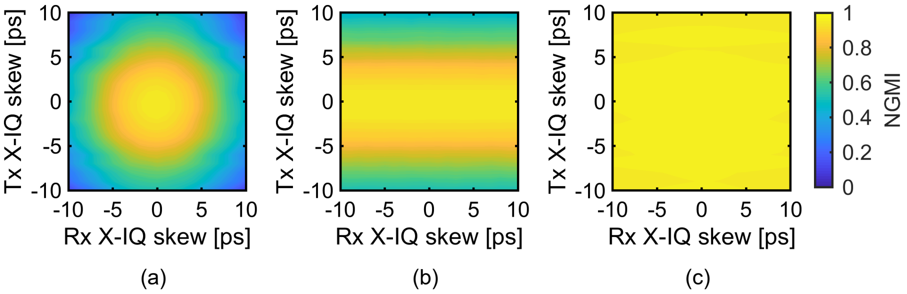


Figure 3.14. Dependence of NGMI on Tx X-IQ skew and Rx X-IQ skew in the cases of (a) 2×2 SL, (b) 4×2 WL, and (c) Multi-layer SL&WL under back-to-back condition.

100-km SMF transmission

Finally, we evaluated the capability of IQ skew compensation after transmission over a 100-km SMF span. The received OSNR was set to 30 dB/0.1 nm. The results of the post-FEC BER while changing X-IQ skew from −10 ps to +10 ps in Tx and Rx are shown in Figs. 3.15(a) and (b), respectively. The corresponding results of the NGMI are shown in Fig. 3.16. Both 2×2 SL and 4×2 WL were unable to compensate for IQ skew in both Tx and Rx, in contrast to the results for the back-to-back condition, due to the existence of accumulated CD and frequency offset. Multi-layer SL&WL was able to compensate for them both, the same as in the back-to-back condition. Figure 3.17 shows the received constellations after impairment compensation with Multi-layer SL&WL for several IQ skew values in Rx. Figure 3.18 shows the dependence of the NGMI on IQ skew in Tx and Rx after 100-km SMF transmission. These results show that the proposed Multi-layer SL&WL could compensate for IQ skew in both Tx and Rx even in the presence of CD accumulation, polarization rotation, and a frequency offset.

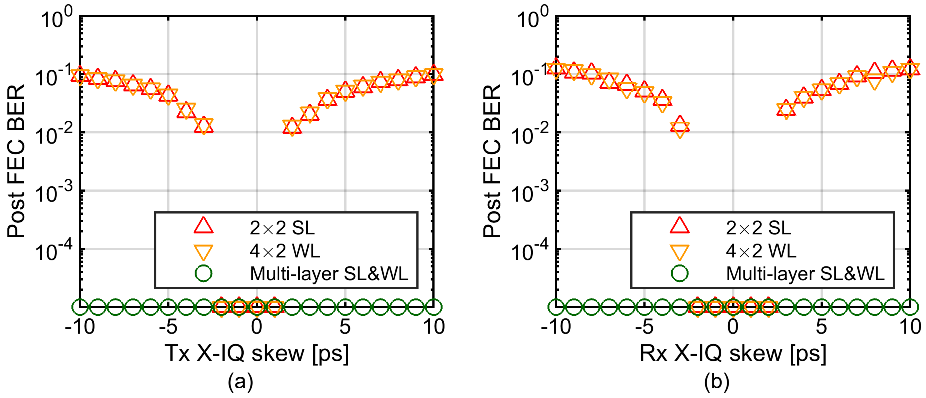


Figure 3.15. Experimental results for post-FEC BER after 100-km SMF transmission with (a) Tx X-IQ skew and (b) Rx X-IQ skew.

3.5 Summary

We developed a multi-layer filter architecture consisting of SL and WL MIMO filters to compensate for relevant impairments including Tx/Rx impairments and adaptive control of their filter coefficients. Taking inspiration from the idea of gradient calculation with back propagation in machine learning with neural networks, we derived an adaptive filter coefficient control algorithm for multi-layer SL and WL FIR filters with SGD based on a loss function composed of the outputs of the last layer. We evaluated the compensation capability of the multi-layer SL&WL filters with adaptive SGD control for IQ skew in both Tx and Rx through simulations with 32-Gbaud PDM-QPSK and an experiment on 32-Gbaud PDM-64QAM transmission over a 100-km SMF span. The results

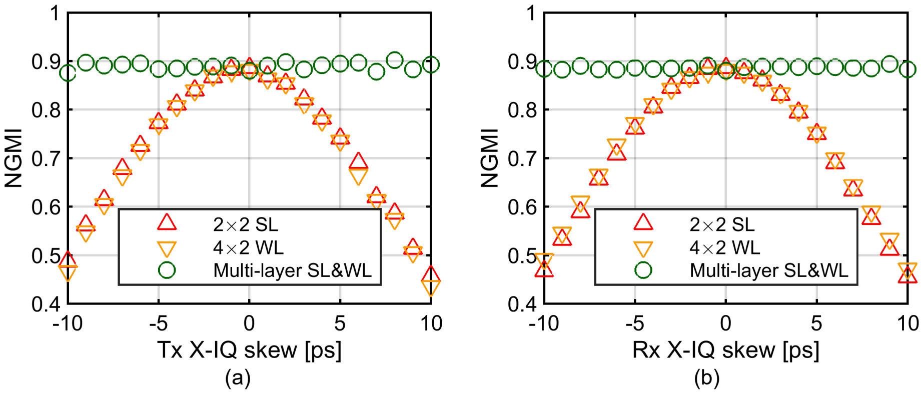


Figure 3.16. Experimental results for NGMI after 100-km SMF transmission with (a) Tx X-IQ skew and (b) Rx X-IQ skew.

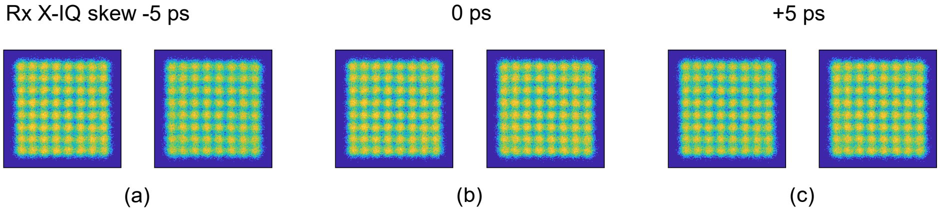


Figure 3.17. Received constellations after impairment compensation DSP with Multi-layer SL&WL after 100-km SMF transmission with Rx X-IQ skew of (a) -5 ps, (b) 0 ps, and (c) +5 ps. The left side shows X polarization and the right side shows Y polarization.

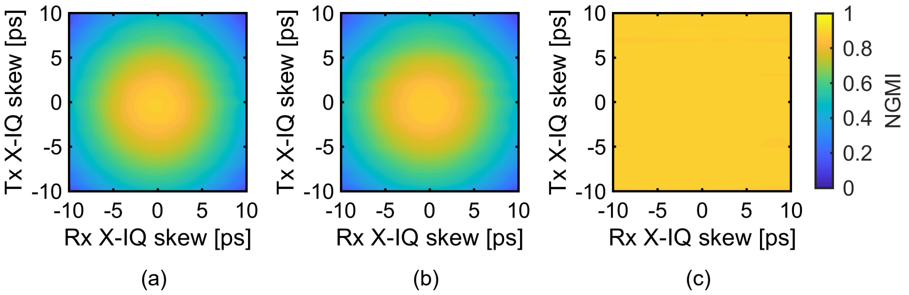


Figure 3.18. Dependence of NGMI on Tx X-IQ skew and Rx X-IQ skew in the cases of (a) 2×2 SL, (b) 4×2 WL, and (c) Multi-layer SL&WL after 100-km SMF transmission.

showed that they could compensate for IQ skew in both Tx and Rx in the presence of CD, polarization rotation, and a frequency offset.

Chapter 4

Simultaneous and individual impairment monitoring using coefficients of adaptive multi-layer strictly linear and widely linear filters

We have confirmed that the multi-layer Sl&WL filter architecture controlled with SGD and back propagation can compensate for Tx and Rx impairments precisely and simultaneously under accumulation CD and a frequency offset in an optical fiber communication system in the previous chapter. This fact implies that the filters in the first and fifth layers in the multi-layer SL&WL filters appropriately serve their expected roles for Rx and Tx impairment compensation while CD or a frequency offset decouples Tx and Rx impairments, although they are controlled simultaneously with SGD rather than independently. The filter coefficients of each filter after the convergence of adaptive control should include information on impairments that are compensated in it. It enables simultaneous and individual monitoring for Tx and Rx impairments.

This chapter investigates simultaneous and individual monitoring for Tx and Rx impairments from the coefficients of corresponding layers in the adaptive multi-layer SL&WL filters. We derive the model relating IQ impairments to the WL filter to compensate for them by considering the order in which IQ skew, IQ imbalance, and IQ phase deviation occur for the Tx and Rx, as we have considered the order of impairments in an optical fiber communication system in the previous chapter. We evaluate the impairment monitor through simulations using 32-Gbaud PDM-QPSK and a 100-km SMF transmission experiment of 32-Gbaud PDM-64QAM. The result that both Tx and Rx impairments can be individually monitored using the WL filter coefficients precisely and simultaneously when multiple impairments exist supports the principle and design of the multi-layer SL&WL filter architecture for impairment compensation in a different perspective.

.

4.1 Background

Optical performance monitoring gives information on the physical layer of an optical communication system about what kind of signal distortion occurs in a system [75]. It provides diagnostics of a system to deal with distortion and valuable information for upper layers of communication systems for network control. As the development of advanced modulation formats and high symbol rate progresses, impairments from imperfections in Tx and Rx devices are becoming a limiting factor [12, 59, 76]. Therefore, compensation for and monitoring Tx and Rx impairments are crucial for future high-speed optical fiber communication systems. For extracting information about impairments that occur in an optical fiber communication system, i.e., for impairment monitoring, several approaches have been investigated depending on the features of the interested impairment [77–81]. In a coherent optical communication system, impairment monitoring from adaptive filter coefficients in the Rx has been investigated [50, 82]. When an adaptive filter works, the filter response consequently reflects the characteristics of the impairment that the filter compensates for, and thus, we can extract information about the impairment from the filter coefficients. This approach is beneficial because dedicated optical and electrical devices for monitoring are not required at all.

Impairment monitoring from adaptive filter coefficients can also be applied to Tx and Rx impairments [9, 54, 57, 58, 83]. However, Tx and Rx impairments are addressed individually in these approaches[1], resulting in deterioration of monitoring accuracy as in the case of impairment compensation discussed in the previous chapter due to imperfect compensation[2].

4.2 Individual IQ impairment monitoring from WL filter coefficients

As well as simultaneous and precise impairment compensation, the adaptive multi-layer SL&WL filter architecture we developed can provide a reasonable solution to this problem. The adaptive multi-layer SL&WL filter can compensate for both Tx and Rx impairments simultaneously and precisely under CD accumulation and a phase/frequency offset. Tx and Rx impairment monitoring from the filter coefficients in the multi-layer SL&WL filters can work if corresponding layers appropriately serve as Tx and Rx impairment compensation individually, though it is not so obvious compared to the case with an individual adaptive filter for Tx or Rx impairment compensation in the previous approach

---

[1] Adaptive filters for Tx and Rx impairment compensation were controlled locally by using their direct inputs and outputs.

[2] Characterizing of individual IQ impairments from coefficients of a one large complex-valued adaptive 8×2 MIMO filter with augmented CD compensation and phase/frequency offset compensation has been investigated recently [84].

since all the filter coefficients in the adaptive multi-layer SL&WL filters are controlled simultaneously with SGD rather than independently.

We first review the models of Tx or Rx impairments including IQ skew, IQ imbalance, and IQ phase deviation by focusing on the order in which they occur in a Tx or Rx. We then derive the relations between IQ impairments and the WL filter responses that compensate for them. We consider the case with IQ modulation and coherent detection.

4.2.1 Models of Tx or Rx impairments

To consider the model of Tx or Rx impairments, it is beneficial to use the IQ component basis. The model for one polarization is enough for our purpose since two orthogonal polarizations are generally handled independently in a Tx and Rx. Considering an input $x(t) = x_{\mathrm{I}}(t) + ix_{\mathrm{Q}}(t)$, where $x_{\mathrm{I}}(t)$ and $x_{\mathrm{Q}}(t)$ are the real-valued I and Q components, and an output $y(t) = y_{\mathrm{I}}(t) + iy_{\mathrm{Q}}(t)$, an arbitrary linear response of a real-valued 2×2 MIMO filter, or an equivalent 2×1 WL filter, is represented in the IQ component basis as Eq. (2.9) in the time-domain or Eq. (2.10) in the frequency-domain.

Tx impairments

We first consider the model of Tx impairments. Figure 4.1(a) shows the configuration of a conventional optical Tx with a nested Mach-Zehnder IQ modulator for coherent optical fiber communication systems [1,56]. The outputs of a DAC modulate a LD source. After mixing with $\pi/2$ phase shift of one side, they construct IQ components. In this configuration, IQ skew and IQ imbalance occur first, then IQ phase deviation occurs. Considering an arbitrary phase rotation of the generated optical signal[3], the frequency response of this model in the IQ component basis is described as

$$\begin{pmatrix} Y_{\mathrm{I}}(\omega) \\ Y_{\mathrm{Q}}(\omega) \end{pmatrix} = H_{\mathrm{Tx}} \begin{pmatrix} X_{\mathrm{I}}(\omega) \\ X_{\mathrm{Q}}(\omega) \end{pmatrix} = H_{\theta} H_{\mathrm{phase}} H_{\mathrm{imb}} H_{\mathrm{skew}} \begin{pmatrix} X_{\mathrm{I}}(\omega) \\ X_{\mathrm{Q}}(\omega) \end{pmatrix}, \tag{4.1}$$

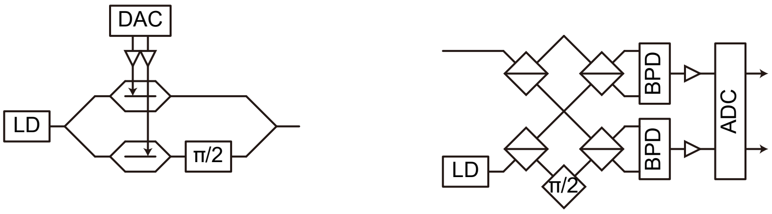
where H_{skew}, H_{imb}, and H_{phase} are the frequency response in the IQ representation with the frequency-domain of IQ skew, IQ imbalance, and IQ phase deviation as shown in Eqs. (2.52), (2.53), and (2.54), respectively. H_{θ} is the frequency response of a phase rotation of an optical signal in the IQ representation. τ, a, and φ in Eqs. (2.52), (2.53), and (2.54) are the amounts of IQ skew, IQ imbalance, IQ phase deviation. θ is the amount of the global phase shift. They can be dependent on frequency, but we omit the frequency dependence for simplicity. H_{skew} and H_{imb} are mutually commutative[4] but not commutative with H_{phase}. We can choose an arbitrary global phase shift as $\theta + \varphi/2$. Then, the frequency response of Tx impairments becomes

$$H_{\mathrm{Tx}} = \begin{pmatrix} \cos(\varphi/2+\theta)(1+a)\exp(i\omega\tau/2) & \sin(\varphi/2-\theta)(1-a)\exp(-i\omega\tau/2) \\ \sin(\varphi/2+\theta)(1+a)\exp(i\omega\tau/2) & \cos(\varphi/2-\theta)(1-a)\exp(-i\omega\tau/2) \end{pmatrix}. \tag{4.2}$$

---

[3] For example, an optical fiber pigtail at the end of the IQ modulator easily adds this phase and usually uncontrollable.

[4] We can understand this based on Fig. 4.1(a). Difference of the lengths of electrical paths of IQ components should provide the same results when it occurs before or after electrical amplification.



(a) (b)

Figure 4.1. Configurations of (a) optical Tx and (b) Rx with coherent detection for one polarization.

Rx impairments

Figure 4.2(b) shows the configuration of a conventional optical Rx with coherent detection [1,56]. The optical signal is mixed with an LO by a 90° optical hybrid, and IQ components are detected with balanced photo-detectors. These signals are then converted into the digital domain by using an ADC for impairment compensation DSP. In contrast to the Tx impairments shown in Fig. 4.2(a), IQ phase deviation comes first, then IQ skew and IQ imbalance occur in this Rx configuration. Based on the model of a 90° optical hybrid [85], IQ phase deviation affects IQ components as $H_{\mathrm{phase}}^{\mathrm{T}}$. Since an arbitrary phase shift to the optical signal is before IQ phase deviation, the frequency response of the Rx model in the IQ component basis is described as

$$\begin{pmatrix} Y_{\mathrm{I}}(\omega) \\ Y_{\mathrm{Q}}(\omega) \end{pmatrix} = H_{\mathrm{Rx}} \begin{pmatrix} X_{\mathrm{I}}(\omega) \\ X_{\mathrm{Q}}(\omega) \end{pmatrix} = H_{\mathrm{imb}} H_{\mathrm{skew}} H_{\mathrm{phase}}^{\mathrm{T}} H_{\theta} \begin{pmatrix} X_{\mathrm{I}}(\omega) \\ X_{\mathrm{Q}}(\omega) \end{pmatrix}. \tag{4.3}$$

With an appropriate choice of an arbitrary phase shift, the frequency response of Rx impairments results in

$$H_{\mathrm{Rx}} = \begin{pmatrix} \cos(\varphi/2-\theta)(1+a)\exp(i\omega\tau/2) & \sin(\varphi/2-\theta)(1+a)\exp(i\omega\tau/2) \\ \sin(\varphi/2+\theta)(1-a)\exp(-i\omega\tau/2) & \cos(\varphi/2+\theta)(1-a)\exp(-i\omega\tau/2) \end{pmatrix}, \tag{4.4}$$

which does not fall in the same response of Tx impairments, though they are similar, because of the order in which IQ impairments occur.

4.2.2 Relation between impairments and WL filter coefficients

We consider again the mutli-layer SL&WL filter architecture for impairment compensation including Tx and Rx impairments consisting of five layers as shown in Fig. 4.2. The five layers are composed of Rx impairment compensation, CD compensation, polarization demultiplexing, carrier recovery, and Tx impairment compensation. All the filters are assumed FIR filters with half-symbol spaced. Coefficients of all the filters can be controlled by SGD with back propagation on the basis of the loss function consisting of the last outputs since all the layers are differentiable in terms of their inputs and their filter coefficients. The first layer consists of two 2×1 WL filters for two polarizations to compensate for Rx impairments. The second layer consists of two 1×1 SL filters for CD compensation, the coefficients of which are treated as static. The third layer consists of a 2×2 SL MIMO filter for polarization demultiplexing. The fourth layer consists of two 1×1 1-tap SL filters to compensate for the phase offset and frequency offset. The fifth layer consists of two 2×1 WL filters to compensate for Tx impairments. The coefficients of the fourth layer are controlled by a phase-locked loop using the last outputs. The coefficients of the first, third, and fifth layers are adaptively controlled by SGD with back propagation.

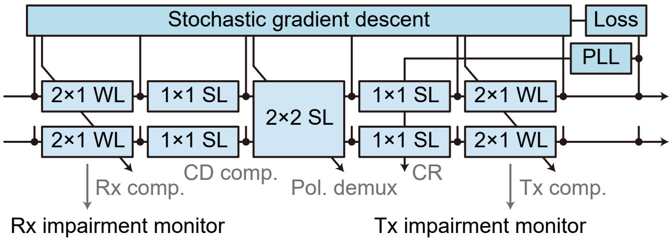


Figure 4.2. Multi-layer SL&WL filters to compensate for all relevant linear impairments including Tx and Rx impairments where coefficients are adaptively controlled by stochastic gradient descent with back propagation from last outputs.

The first and fifth WL filter layers are expected to serve as Rx and Tx impairment compensation in this multi-layer filters. Thus, their filter coefficients after adaptive control should contain information about Rx and Tx impairments if the adaptive filters work well. We now derive the relations between Tx and Rx impairments and the WL filter responses that compensate for them. For considering IQ impairments including IQ skew, IQ imbalance, and IQ phase deviation, using the IQ basis representation is advantageous. The complex-valued linear response functions of a 2×1 WL filter of $\boldsymbol{h}$ and $\boldsymbol{h}_*$, with which the input-output relation is described as Eq. (2.12), are related to the 2×2 real-valued linear response functions of $\boldsymbol{h}_{ij}$ $(i, j = \mathrm{I}, \mathrm{Q})$ as Eq. (2.19). Using this relation, we can obtain the linear responses of a WL filter in the IQ basis representation and their corresponding frequency responses.

We represent the frequency response of a 2×1 WL filter in the IQ representation as

$$W = \begin{pmatrix} W_{\mathrm{II}} & W_{\mathrm{IQ}} \\ W_{\mathrm{QI}} & W_{\mathrm{QQ}} \end{pmatrix}. \tag{4.5}$$

Ignoring noise contribution, the frequency responses of the first WL filter after the convergence of adaptive control in the IQ basis representation $W^{[1]}$ are expected to satisfy $W^{[1]} = gH_{\mathrm{Rx}}^{-1}$ with a certain complex-valued gain g[5]. Assuming this relation, extracting IQ skew τ, IQ imbalance a, and IQ phase deviation

---

[5] This cannot exactly be satisfied since adaptive control is executed on the basis of a certain

φ occurring in an Rx from the coefficients of the first WL filter comes down to nonlinear simultaneous equations. The solutions are provided in a heuristic manner as follows. We consider these four values, which are independent of the global phase shift of θ, as

$$\begin{aligned} A^{[1]} &= W_{\mathrm{II}}^{[1]} W_{\mathrm{QQ}}^{[1]} - W_{\mathrm{QI}}^{[1]} W_{\mathrm{IQ}}^{[1]}, && (4.6)\\ B^{[1]} &= W_{\mathrm{II}}^{[1]} W_{\mathrm{QQ}}^{[1]*} - W_{\mathrm{QI}}^{[1]} W_{\mathrm{IQ}}^{[1]*}, && (4.7)\\ C^{[1]} &= W_{\mathrm{QI}}^{[1]} W_{\mathrm{QQ}}^{[1]} + W_{\mathrm{II}}^{[1]} W_{\mathrm{IQ}}^{[1]}, && (4.8)\\ D^{[1]} &= \sqrt{\frac{|W_{\mathrm{II}}^{[1]}|^2 + |W_{\mathrm{QI}}^{[1]}|^2}{|W_{\mathrm{IQ}}^{[1]}|^2 + |W_{\mathrm{QQ}}^{[1]}|^2}}. && (4.9) \end{aligned}$$

By using these values, we can obtain

$$a_{\mathrm{Rx}} = \frac{1 - D^{[1]}}{1 + D^{[1]}}, \tag{4.10}$$

and

$$\varphi_{\mathrm{Rx}} = \tan^{-1}\left(-\frac{C^{[1]}}{A^{[1]}}\right). \tag{4.11}$$

If IQ phase deviation φ is small and the variation of IQ skew τ over frequency is slow,

$$\tau_{\mathrm{Rx}} = -\frac{d}{d\omega}\arg(B^{[1]}). \tag{4.12}$$

Similarly, assuming $W^{[5]} = gH_{\mathrm{Tx}}^{-1}$ for Tx impairments, we obtain

$$\begin{aligned} A^{[5]} &= W_{\mathrm{II}}^{[5]} W_{\mathrm{QQ}}^{[5]} - W_{\mathrm{IQ}}^{[5]} W_{\mathrm{QI}}^{[5]}, && (4.13)\\ B^{[5]} &= W_{\mathrm{II}}^{[5]} W_{\mathrm{QQ}}^{[5]*} - W_{\mathrm{IQ}}^{[5]} W_{\mathrm{QI}}^{[5]*}, && (4.14)\\ C^{[5]} &= W_{\mathrm{IQ}}^{[5]} W_{\mathrm{QQ}}^{[5]} + W_{\mathrm{II}}^{[5]} W_{\mathrm{QI}}^{[5]}, && (4.15)\\ D^{[5]} &= \sqrt{\frac{|W_{\mathrm{II}}^{[5]}|^2 + |W_{\mathrm{IQ}}^{[5]}|^2}{|W_{\mathrm{QI}}^{[5]}|^2 + |W_{\mathrm{QQ}}^{[5]}|^2}}, && (4.16) \end{aligned}$$

and

$$a_{\mathrm{Tx}} = \frac{1 - D^{[5]}}{1 + D^{[5]}}, \tag{4.17}$$

$$\varphi_{\mathrm{Tx}} = \tan^{-1}\left(-\frac{C^{[5]}}{A^{[5]}}\right), \tag{4.18}$$

$$\tau_{\mathrm{Tx}} = -\frac{d}{d\omega}\arg(B^{[5]}). \tag{4.19}$$

Using there relations, we can monitor each of Tx and Rx impairments simultaneously and individually from the coefficients of the WL filters in the multi-layer SL&WL filters if the adaptive control by SGD with back propagation goes well.

---

criterion e.g., mean squared error under noise. We assume it is satisfied approximately since the SNR of the received signal should be high enough to accomplish communication at least.

4.3 Evaluation in simulation

We first evaluated the monitoring method for individual Tx and Rx impairments from the coefficients of the multi-layer SL&WL filters through numerical simulations. We used a 32-Gbaud PDM-QPSK signal with a root raised cosine filter, where a roll off factor of 0.1, was transmitted over a 100-km SMF and received with coherent detection. On the basis of the above impairment model described in the previous section, IQ skew, IQ imbalance, and IQ phase deviation were imposed in the Tx and/or Rx. Only CD was given in SMF transmission. We assumed no laser phase noise and frequency offset at first. The received optical signal-to-noise ratio (OSNR) was set to 30 dB/0.1 nm by adding white Gaussian noise. DSP was applied to the received signals sampled by two-fold oversampling after coherent detection.

In DSP, the received signals were first normalized as complex-valued signals for two orthogonal polarizations then resampled to two-fold oversampling based of the timing error [86][6]. After matched filtering, the above-mentioned five-layer adaptive multi-layer SL&WL filters were applied. The tap lengths of the filters were 61 taps for CD compensation (the second layer), 21 taps for polarization demultiplexing (the third layer), and five taps to compensate for Tx/Rx impairments (the first and fifth layers). The filter coefficients of the second layer were set to compensate for CD and did not updated. The filter coefficients at the center tap of the main diagonal position in the first, third, and fifth layers were initialized as one and the rest were set to zero. To monitor IQ impairments in a Tx and Rx, the transmitted IQ components should be identified. Therefore, the constant modulus algorithm (CMA) and decision directed LMS are not suitable since they include a certain phase ambiguity that results in the exchange of IQ components. The loss function based on data-aided LMS was thus used for adaptive control with SGD to avoid the uncertainty of the IQ components. In this simulation, the transmitted pattern was regarded as a training sequence for simplicity. The step sizes of the update were 10^{-4} for the first and fifth layers and 10^{-3} for the third layer.

After convergence of the adaptive filter control, the WL filter coefficients of the first layer were converted to the IQ basis representation then converted in the frequency domain by the fast Fourier transform (FFT). The amounts of IQ skew, IQ imbalance, and IQ phase deviation in the Rx were estimated using the equations described in the previous section. The values of the frequency components of 0 Hz were evaluated for IQ imbalance and IQ phase deviation. For IQ skew estimation, derivative was replaced with the finite difference between the components of 0 Hz and $\Delta\omega/(2\pi) = 12.8$ GHz. IQ skew, IQ imbalance, and IQ phase deviation in the Tx were estimated in a similar manner with the WL filter coefficients of the fifth layer.

Individual IQ impairment

We first evaluated the performance of compensation and impairment monitoring under the condition where only one of the impairments in the Tx and Rx was imposed. IQ skew, IQ imbalance, and IQ phase deviation of X polarization in the Tx or Rx were swept from -5 to +5 ps, -0.5 to 0.5, -10° to +10°, respectively.

---

[6] The timing error was calculated with CD compensation on the resampled output signals and the timing offset in resampling of the input signals before CD compensation was optimized.

Figure 4.3 shows the results of the averaged error vector magnitude (EVM) over two polarizations after the multi-layer SL&WL filters. As a reference, two other DSP architectures compared in the previous chapter were also evaluated. The results where only CD compensation, polarization demultiplexing by a 2×2 MIMO filter, and carrier recovery were carried out are shown in this figure (referred to as 2×2 SL). In addition, the results where a 4×2 WL MIMO filter was used instead of a 2×2 MIMO filter of 2×2 SL are shown (referred to as 4×2 WL). Regarding IQ skew and IQ phase deviation in the Tx or Rx, an almost constant EVM was obtained after using the multi-layer SL&WL filters while EVM degraded in the case of 2×2 SL as the impairment increased. EVM after the multi-layer SL&WL filters became worse with IQ imbalance in the Tx since it distorted added Gaussian noise, though compensation still worked. These results indicate that the multi-layer SL&WL filters could compensate for each of Tx and Rx impairments. In the case of 4×2 WL, only Tx impairments could be compensated since CD was included and a frequency offset and phase noise were not included.

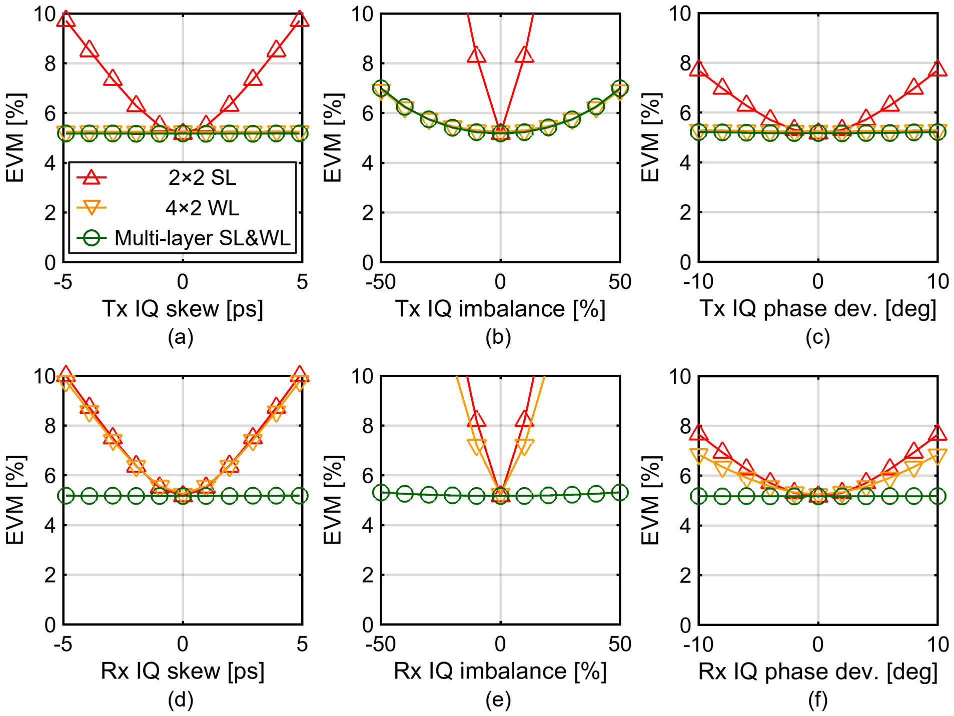


Figure 4.3. Simulation results of EVM with (a) Tx IQ skew, (b) Tx IQ imbalance, (c) Tx IQ phase deviation, (d) Rx IQ skew, (e) Rx IQ imbalance, and (f) Rx IQ phase deviation.

The results of the impairment estimation for IQ skew, IQ imbalance, and IQ phase deviation in the Tx and Rx by using the converged WL filter coefficients of the multi-layer SL&WL filters are shown in Fig. 4.4. We can confirm that the imposed impairments and estimated ones agreed well.

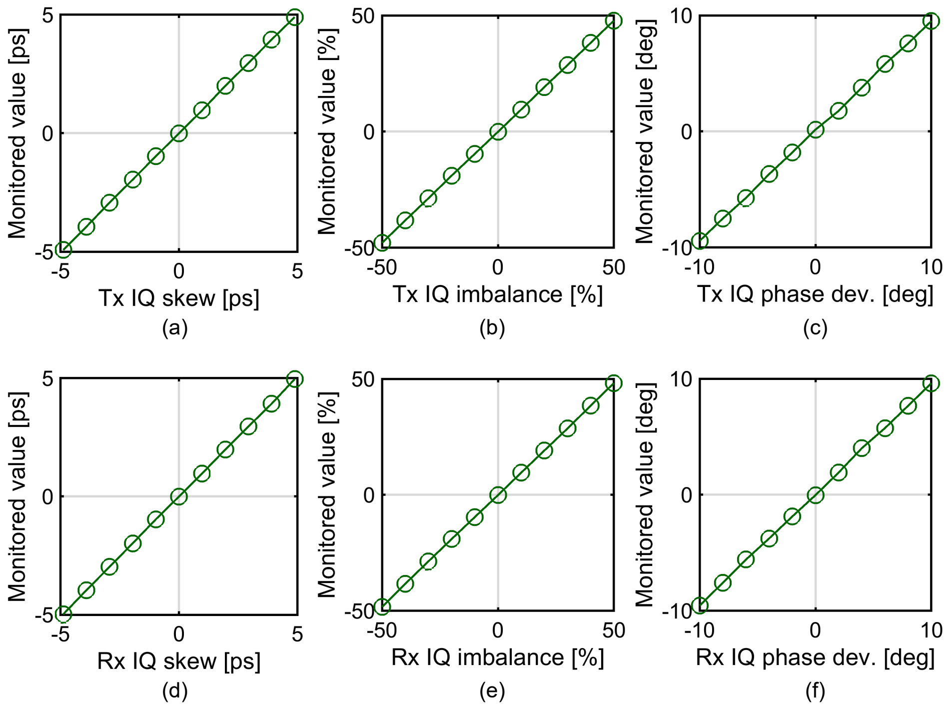


Figure 4.4. Simulation results of impairment monitoring for (a) Tx IQ skew, (b) Tx IQ imbalance, (c) Tx IQ phase deviation, (d) Rx IQ skew, (e) Rx IQ imbalance, and (f) Rx IQ phase deviation.

Simultaneous random IQ impairments

We then evaluated the impairment monitoring with our method under the condition where multiple impairments were imposed simultaneously. IQ skew, IQ imbalance, and IQ phase deviation of X and Y polarization in the Tx and Rx were random values of a zero-mean Gaussian distribution with a standard deviation of 3 ps (10% of the symbol duration), 0.1, and 5°, respectively. Random rotation of the polarization state was also added in SMF transmission in addition to CD. The LO phase noise of the linewidth of 100 kHz was included. Figure 4.5 shows the results for monitoring the impairments of X polarization with 1000 random realizations of multiple impairments. The monitored values are plotted on the left axis, and the monitored errors from the imposed amounts are plotted on the right axis. Similar results were obtained for Y polarization as shown in Fig. 4.6. Similar to the results where only one of the impairments was imposed, as shown in Fig. 4.4, the imposed impairments and estimated values agreed well in this case where multiple impairments were imposed simultaneously. In accordance with the monitored error, the monitored values were biased due to the nonlinear relations to the WL filter coefficients, while they were small within about ±0.2 ps, ±0.02, and ±1° for IQ skew, IQ imbalance, and IQ phase deviation, respectively. The results indicate that the impairments in both Tx and Rx could be individually monitored from the WL filter coefficients precisely and simultaneously when multiple impairments existed.

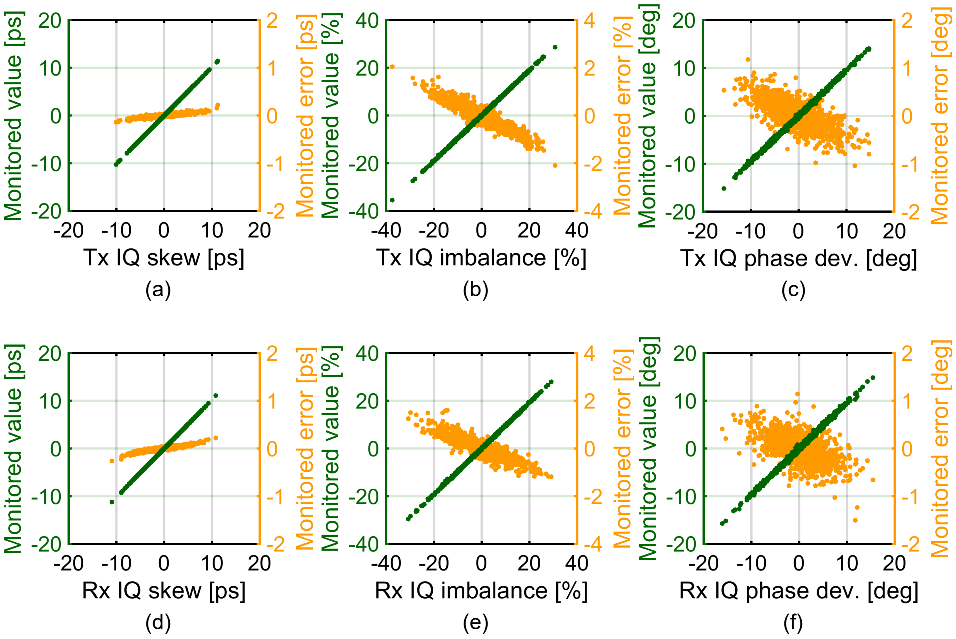


Figure 4.5. Simulation results of impairment monitoring under simultaneous random multiple impairment conditions for (a) Tx IQ skew, (b) Tx IQ imbalance, (c) Tx IQ phase deviation, (d) Rx IQ skew, (e) Rx IQ imbalance, and (f) Rx IQ phase deviation of X polarization. Results for 1000 random realizations of multiple impairments are plotted in each figure.

4.4 Evaluation in transmission experiment

We evaluated the monitoring method for individual impairments in a Tx and Rx from the coefficients of the multi-layer SL&WL filters in a transmission experiment of 32-Gbaud PDM-64QAM over a 100-km SMF span.

Experimental setup

A schematic diagram of the experimental setup is shown in Fig. 4.7, which is a similar configuration to that in the previous chapter. The 32-Gbaud PDM-64QAM was generated by modulating a laser source at the frequency of 193.3 THz having a linewidth of about 100 kHz with waveforms from a four-channel

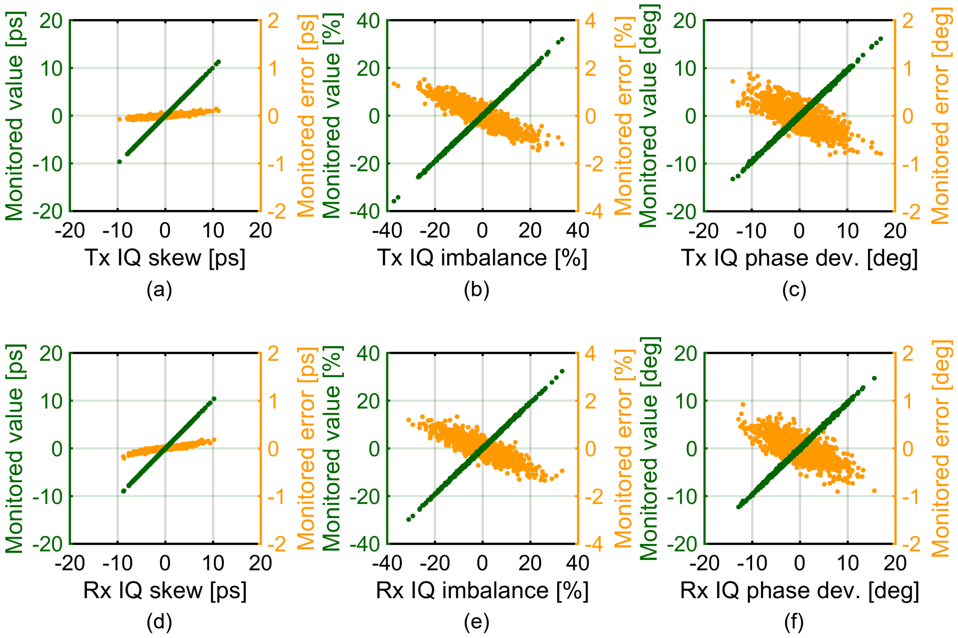


Figure 4.6. Simulation results of impairment monitoring under simultaneous random multiple impairment conditions for (a) Tx IQ skew, (b) Tx IQ imbalance, (c) Tx IQ phase deviation, (d) Rx IQ skew, (e) Rx IQ imbalance, and (f) Rx IQ phase deviation of Y polarization. Results for 1000 random realizations of multiple impairments are plotted in each figure.

DAC at the sampling rate of 92 GS/s with a vertical resolution of eight bits. Transmitted data were 16 frames of low-density parity-check code for DVB-S2 with a frame length of 64,800 and code rate of 0.8 while random bits were loaded to their payload. These data were mapped to PDM-64QAM with Gray mapping. Pilot symbols of QPSK were inserted every 15 symbols to execute a pilot-based data-aided LMS. Due to the restriction of the data length that can be handled with the DAC used in the experiment, an overhead of QPSK symbols was inserted, which resulted in about 8% overhead in total with the pilot. The inserted QPSK symbols were set to any of the outer symbol points of 64QAM. The data were upsampled to two-fold oversampling, and a root-raised cosine filter with a roll-off factor of 0.1 was used. After using a pre-compensation filter to calibrate the frequency characteristic in Tx devices, the data for the IQ of each polarization were resampled to 92 GS/s to be generated with the DAC.

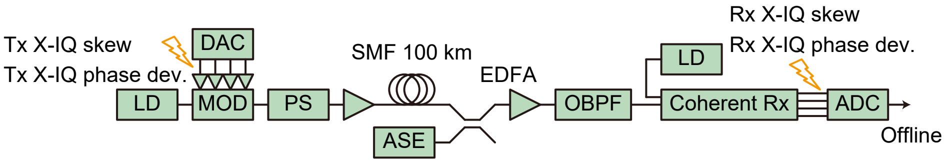


Figure 4.7. Experimental setup of 32-Gbaud PDM-64QAM transmission over 100-km SMF span.

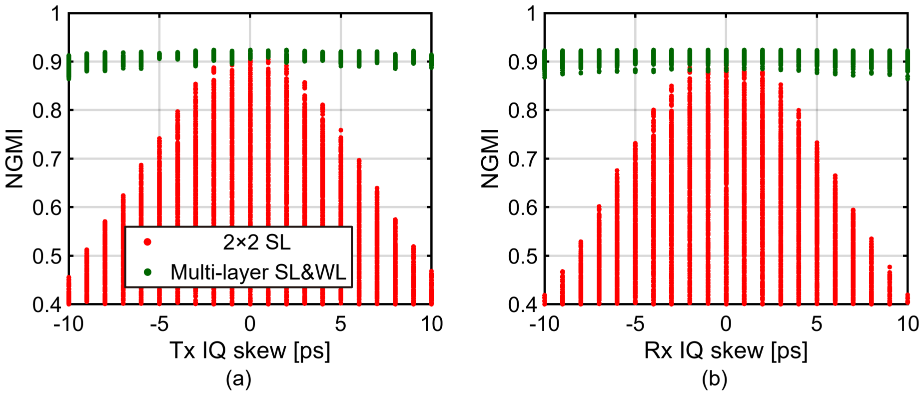
A 32-Gbaud PDM-64QAM signal was transmitted to a 100-km SMF after low-speed polarization scrambling with a speed of $10\times2\pi$ rad/s. The span input optical power was set to 0 dBm. After 100-km SMF transmission, amplified spontaneous emission was added to set the OSNR to 30 dB/0.1 nm. The received optical signal was amplified using an EDFA and filtered using an optical bandpass filter having a 3-dB bandwidth of 50 GHz. The optical signal was received using a polarization-diversity coherent receiver with a LO source having a linewidth of about 100 kHz. The laser sources for the signal and LO were free running. The FO was about 60 MHz on average, and its standard deviation was about 90 MHz. The four outputs of the coherent receiver were sampled using an oscilloscope as the ADC at the sampling rate of 80 GS/s with a vertical resolution of eight bits.

The Tx and Rx impairments were emulated digitally as arbitrary impairments were difficult to impose in the analog domain. In this experiment, we mainly focused on IQ skew in the Tx and Rx. IQ skew of X polarization from -10 to +10 ps in steps of 1 ps were imposed to the signal generated from the DAC and to those obtained with the ADC. For the condition of multiple impairments, we also evaluated the cases in which the IQ phase deviation of X polarization of 0° and ±10° was digitally emulated in the Tx and Rx. Strictly speaking, the IQ phase deviation imposed in this manner did not fully match the impairment model described in the previous section if there remained IQ skew in the Tx and Rx devices, though they were calibrated as much as possible. Since the average amplitudes of IQ components are relatively easy to access in current Tx/Rx transceivers, we did not include IQ imbalance explicitly in this experiment. For each combination of impairments, the received waveforms were acquired three times.

DSP was carried out offline. The impairment compensation DSP was almost the same as that used in the above simulation, though pilot-based data-aided LMS was applied in this experiment. Pilot-based DSP was achieved as follows. After CD compensation, the pilot position was detected in accordance with the difference in the averaged powers between the 64QAM signal and QPSK pilot. Polarization demultipexing was then carried out using CMA for the pilot. Timing alignment of the received signal and pilot sequence was achieved using their correlation. With the timing alignment, the main DSP with the multi-layer SL&WL filters was applied to the signal before CD compensation again. The configuration of the multi-layer SL&WL filters were the same as in the simulation. The filter coefficients of the first, third, and fifth layers were adaptively updated using the QPSK pilot with data-aided LMS. The step sizes of the update were 10^{-3} for the first and fifth layers, and 10^{-2} for the third layer. After the adaptive multi-layer SL&WL filters, the Gram-Schmidt orthogonalization procedure was carried out, then the pilot and overhead were removed to evaluate the NGMI as a performance indicator [74]. The amounts of IQ skew, IQ imbalance, and IQ phase deviation in the Tx and Rx were estimated using the WL filter coefficients of the first and the fifth layers as well.

Transmission performance

Figure 4.8 shows the results of NGMI for all the $21^2 \times 3^2$ combinations of IQ skew and IQ phase deviation in the Tx and Rx and 3 acquisitions, simultaneously. The results are arranged as the function of IQ skew in the Tx in Fig. 4.8(a) and as that in the Rx in Fig. 4.8(b). The results with 2×2 SL are also plotted as a reference. With the multi-layer SL&WL filters, a stable NGMI of about 0.9 was obtained regardless of the impairments.



(a)
(b)

Figure 4.8. Experimental results of NGMI with (a) Tx IQ skew and (b) Rx IQ skew. At each Tx (Rx) IQ skew, results of Rx (Tx) IQ skew from − 10 to +10 ps and Tx/Rx IQ phase deviation of 0° and ±10° are plotted.

We compared the convergence speed of the adaptive multi-layer SL&WL filters when multiple impairments existed. A learning curve of the loss of data-aided LMS at the pilot symbols, that is, the training loss is the indicator of the convergence performance. Figure 4.9 shows the time development of the training losses obtained in the 100-km SMF transmission experiment of 32-Gbaud PDM-64QAM. Even when multiple impairments existed, the convergence speed was little affected and the losses are almost converged after 2000 symbols.

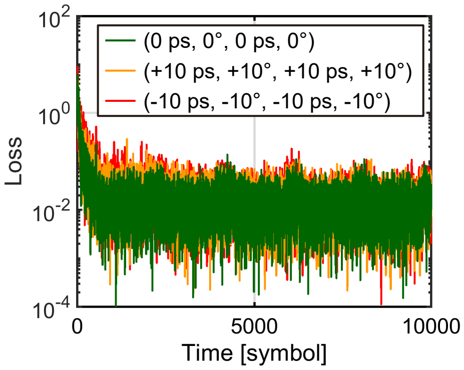


Figure 4.9. Experimental results of time development of loss function. (Tx IQ skew, Tx IQ phase deviation, Rx IQ skew, Rx IQ phase deviation) are (0 ps, 0°, 0 ps, 0°), (+10 ps, +10°, +10 ps, +10°), and (− 10 ps, − 10°, − 10 ps, − 10°).

Impairment monitoring

Figure 4.10 shows the results of the impairment monitoring for IQ skew of X polarization in the Tx and Rx. Similarly, the results for all combinations of IQ skew and IQ phase deviation in the Tx and Rx are plotted simultaneously, arranged as a function of IQ skew in Tx (Fig. 4.10(a)) and Rx (Fig. 4.10(b)), respectively. The monitored values are plotted on the left axis and the monitored errors from the emulated values are plotted on the right axis. For IQ skew emulated in both the Tx and Rx, the estimated values corresponded to the emulated ones. The monitored errors were within ±0.5 ps. The results confirm that the impairments in both Tx and Rx could be individually monitored precisely and simultaneously from the WL filter coefficients in the multi-layer SL&WL filters adaptively controlled by SGD with back propagation. Since this monitoring scheme relies on pilot-based adaptive equalization, it can be used with other modulation formats including probabilistic constellation shaping straightforwardly if pilot-based signal processing is adopted.

4.5 Summary

We proposed a monitoring method for individual impairments in a Tx and Rx including IQ skew, IQ imbalance, and IQ phase deviation on the basis of the fil-

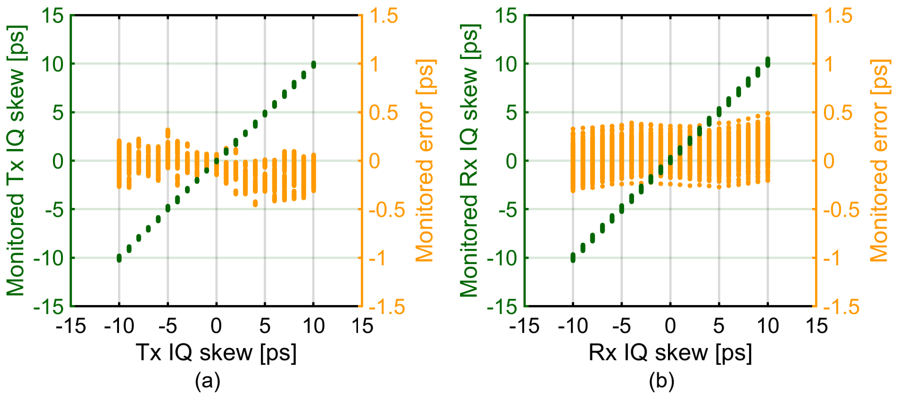


Figure 4.10. Experimental results of impairment monitoring for (a) Tx IQ skew and (b) Rx IQ skew. At each Tx (Rx) IQ skew, results of Rx (Tx) IQ skew from − 10 to +10 ps and Tx/Rx IQ phase deviation of 0° and ±10° are plotted.

ter coefficients of our adaptive multi-layer SL&WL filter architecture controlled by SGD with back propagation. Considering the order in which IQ impairments occur in a Tx or Rx, the model relating IQ skew, IQ imbalance, and IQ phase deviation to the corresponding WL filter responses in the multi-layer filter architecture were obtained. The method was evaluated through simulations using 32-Gbaud PDM-QPSK and a 100-km SMF transmission experiment of 32-Gbaud PDM-64QAM. The results indicate that IQ skew, IQ imbalance, and IQ phase deviation in both Tx and Rx could be individually monitored from the WL filter coefficients precisely and simultaneously when multiple impairments existed.

Chapter 5

Adaptive multi-layer filters incorporated with Volterra filters for impairment compensation of transmitter and receiver nonlinearity

Advanced modulation formats with high spectral efficiency are susceptible to not only linear impairments but also nonlinear characteristics due to device imperfection in a Tx and Rx. Since nonlinear processes are not mutually commutative with any linear and nonlinear processes in general, we have to consider the order in which impairments occur and they are compensated as we consider it for linear MIMO processes in the previous chapters. The multi-layer filter structure gives a reasonable approach to dealing with it. In this chapter, we extend the adaptive multi-layer filter architecture by incorporating differentiable nonlinear layers to compensate for Tx and Rx nonlinearity in optical fiber communication systems. We introduce Volterra filters as nonlinear layers in the adaptive multi-layer filters. The coefficients of the Volterra filters are adaptively controlled by SGD and gradient calculation with back propagation. We evaluated the performance of the adaptive multi-layer filters including Volterra filter layers through simulations with a simple model and experiments where more realistic Tx and Rx nonlinearity was induced by tuning the output amplitude of electronic amplifiers. The adaptive multi-layer filters are used in receiver-side signal processing for the transmission of a 23 Gbaud PDM-64QAM signal over one span of a 100-km SMF. The results demonstrate that the architecture can compensate for the nonlinearity that occurs in both Tx and Rx simultaneously and effectively under the accumulation of other impairments such as CD.

5.1 Background

Nonlinear impairments in a Tx and Rx in optical fiber communication systems are mainly caused by DACs, electronic driver amplifiers, and a Mach-Zehnder modulator in the Tx, as well as electronic TIAs and ADCs in the Rx. These impairments are becoming non-negligible, especially for signals with a high symbol rate where high frequency devices are used [13, 87–89]. Characteristics of nonlinear impairments depend on the components used in a Tx and Rx, which are usually unknown beforehand. Thus, an adaptive approach or learning is required to deal with Tx and Rx impairments.

To compensate for nonlinear impairments that occur mainly in a Tx, digital pre-distortion in the Tx side has been investigated on the basis of Volterra filters [90–93] and neural networks [94–96]. These pre-distortion approaches enable adaptive equalization by using a signal with a high SNR without the effect from other impairments that occur in a fiber transmission; however, they can only resolve Tx nonlinearity. Different approaches are required to compensate for nonlinear impairments that occur in an Rx. Moreover, other effects such as CD accumulate in a signal though fiber propagation. Nonlinear impairments are not mutually commutative with other effects, so compensating for fiber nonlinearity uses a split-step back propagation based on the nonlinear Schrödinger equation [97]. Therefore, to compensate for nonlinear impairments together with other impairments, the order in which all the relevant impairments occur should be considered unless one lumped adaptive nonlinear filter is used. However, a conventional DSP uses a block-wise compensation to effectively deal with various impairments that have different causes and models [5][1]. From this point of view, mutual non-commutativity of nonlinear impairments that occur in a Tx and Rx with other impairments has not been resolved in these previous approaches, preventing compensation of both Tx and Rx nonlinearity at the same time. Recently, combination of pre-distortion at the Tx side and adaptive nonlinear equalization at the Rx side has been reported [98], where a nonlinear equalizer for Rx nonlinearity compensation is positioned at the first of impairment compensation blocks and a nonlinear equalizer for Tx nonlinearity compensation is positioned at the last. Four real-valued nonlinear filters were used for both the Rx and Tx nonlinear equalizers and no impairment compensation blocks that have IQ cross terms were included in this DSP. It is reasonable to use four real-valued nonlinear filters since Tx and Rx nonlinearity usually affects IQ components independently in coherent optical transmission systems. Whereas, the absence of IQ cross terms prevents compensation of IQ phase deviation. If nonlinear equalizers that have IQ cross terms are used, this problem will be resolved, though this straightforward approach increases the number of parameters of nonlinear equalizers greatly, resulting in high computational complexity. Consequently, the features of Tx and Rx nonlinearity in optical fiber communication systems, including mutual non-commutativity, were not treated sufficiently in these previous approaches.

---

[1] It is computationally inefficient to compensate for all these linear and nonlinear impairments that occur in an optical fiber communication system with one large nonlinear filter, since it should have a large degree of freedom, i.e. tremendous cross terms and nonlinear terms with long memory.

5.2 Adaptive multi-layer filters incorporated with Volterra filters

We extended the adaptive multi-layer filter architecture by incorporating nonlinear filters to compensate for both Tx and Rx nonlinearity when other impairments such as CD coexist. Volterra filters and neural networks are both nonlinear functions and back propagation can be applied to both of them. A deep neural network (DNN) slightly outperforms in compensating nonlinearity with memory effects when nonlinear compensation is performed after conventional linear impairment compensation [33]. From the view point of commutativity of impairments, this previous work is regarded as a nonlinearity compensation in a Tx. Although DNNs have the ability to approximate a complicated nonlinear function, random initialization of parameters is usually required before learning [63], resulting completely random outputs at an initial phase. Regarding impairment compensation in optical fiber communications, dominant sources to prevent demodulation are linear effects that occur in fiber propagation such as CD, though Tx and Rx nonlinearity cannot be ignored. Therefore, initializing nonlinear filters that compensate for Tx and Rx nonlinearity as a certain linear or even an identity function instead of a random function can help convergence at the beginning of adaptive control. In the case of the Volterra filter, which is easily initialized as a linear filter, an optimum nonlinear function can be smoothly and steadily obtained by adaptive control from the initial state. Here, we introduced Volterra filters into the adaptive multi-layer filters. Considering the order in which all the relevant impairments occur, the multi-layer filters consist of SL and WL filter layers to compensate for relevant linear impairments, and the two Volterra filter layers, each of which works as to compensate for nonlinearity that occurs in an Rx and Tx, respectively, are appropriately positioned in the multi-layer filters. The coefficients including the Volterra filter layers were adaptively controlled by a gradient calculation with back propagation and SGD. In this multi-layer filter architecture including Volterra filter layers, a Volterra filter itself compensates only for Rx or Tx nonlinearity and is not required to compensate for any other effects and their interaction with nonlinearity, which expands a temporal spread. Thus, the Volterra filter layers in the multi-layer filters can be implemented with short memory taps, though the number of coefficients and computational complexity of a Volterra filter increase drastically with the increase in length of the memory taps [99].

We first review the nonlinear impairments that occur in optical fiber communication systems with coherent detection. We then introduce the multi-layer filter architecture including Volterra filter layers in an appropriate order and its adaptive control. The coefficients are updated by a gradient calculation with back propagation and SGD to minimize a loss function that is composed of the last layer outputs. Although an adaptive Volterra filter is well-known [100], adaptive control of the filter coefficients with its direct input and output is insufficient when incorporating it into the adaptive multi-layer filters. We derive the back propagation of the Volterra filter layers in the multi-layer filters, in other words, calculating the gradients of a loss in terms of filter coefficients and inputs, when gradients in terms of filter outputs are given, to update the coefficients in all the layers including the Volterra filter ones.

We consider a schematic diagram of a coherent WDM transmission system and Tx and Rx as previously shown in Figs. 2.2 and 2.3. On the Tx side shown in Fig. 2.2(a), some nonlinearity occurs in the DAC and electric driver amplifiers for IQ components [101]. A Mach-Zehnder modulator also has nonlinear sinusoidal characteristics [85]. On the Rx side, TIAs in a coherent receiver for IQ components and ADC induce nonlinearity. As we discussed in the previous chapters, linear IQ impairments such as IQ skew, IQ imbalance, and IQ phase deviation also occur in the Tx and Rx. In a fiber propagation between the Tx and Rx, CD and PMD accumulate in the signal. We ignore the fiber Kerr nonlinearity for simplicity[2]. Linear processes including Tx and Rx linear impairments, CD, and PMD can obviously change the intensity profile of a signal, and thus nonlinear processes are not commutative with them.

Figure 5.1 shows the adaptive multi-layer filter architecture where two Volterra filter layers are positioned to compensate for Rx and Tx nonlinearity considering the order of impairments. It is composed of eight layers comprising FIR filters and Volterra filters. We consider here that both FIR and Volterra filters operate at half-symbol-spaced. The first layer has Volterra filters to compensate for Rx nonlinearity. Although complex-valued Volterra filters exist [104], a nonlinear filter does not have to compensate for any interaction between nonlinear impairment and other effects in this multi-layer filter architecture. Considering the architecture of transmission systems with coherent detection as shown in Fig. 2.3, it is valid to assume that Rx nonlinearity imposed to I and Q components in terms of the Rx output independently by Rx electronic amplifiers. Nonlinearity imposed by ADCs also affects IQ components in terms of the Rx output independently. Thus, there are no nonlinear mixing between IQ components in terms of the Rx output at the input of the first layer, and the first layer requires no IQ cross terms to compensate for Rx nonlinearity in this case. Therefore, four Volterra filters with real-valued input and real-valued coefficients are used for the corresponding IQ components. The second layer has two 2×1 WL filters for two polarizations to compensate for Rx linear impairments. Focusing on IQ mixing that occurs in an Rx, which cannot be compensated by the first layer having four real-valued Volterra filters, it occurs at a coherent receiver as a phase deviation, which is followed by Rx electronic amplifiers. Thus, Rx nonlinearity compensation is followed by Rx linear impairment compensation. This Rx nonlinear and linear compensation chain can mitigate complexity of nonlinear filters since no IQ cross terms are required for nonlinear filters, whereas IQ mixing can be compensated by linear filters. The third layer has two 1×1 SL filters to compensate for CD, whose coefficients are static and determined by the amount of accumulated CD. The fourth layer has a 2×2 SL MIMO filter to perform polarization demultiplexing and PMD compensation, The fifth layer has two 1×1 SL 1-tap filter to perform carrier recovery, whose coefficients, or amounts of phase compensation, are controlled by a digital PLL with the last layer outputs. The sixth layer has two 2×1 WL filters to compensate for Tx linear impairments. The seventh layer has Volterra filters to compensate for Tx nonlinearity. Similar to Rx nonlinearity, it is valid to assume that Tx nonlinearity imposed to I and Q components in terms of the Tx input independently by Tx electronic amplifiers. Nonlinear characteristics of a Mach-Zehnder modular, which affect IQ components in terms of the Tx input independently, follow it and then a phase

---

[2] Actually. much effort has been towards to compensation of fiber nonlinearity [7,17,48,102, 103], rather than Tx and Rx nonlinearity in the research filed of optical fiber communication.

deviation, which causes IQ mixing, follows. Therefore, as with the first layer, four real-valued Volterra filters are used for the seventh layer and they follow Tx linear impairment compensation. At the input of the seventh layer, all mixing between IQ components in terms of the Tx input is compensated through previous six layers. The eighth layer has four real-valued linear filters for the IQ components that work as a matched filter. Narrow bandwidth filtering like a matched filter can change a signal intensity profile, and is not commutative with nonlinear effects. The matched filter is thus incorporated into the last of the multi-layer filters. The coefficients of the first, second, fourth, sixth, and seventh layers are adaptively updated to minimize a loss that is composed of the last layer outputs by a gradient calculation with back propagation from the last layer and SGD. By this adaptive control, the coefficients of the Volterra filters are converged without involving matrix inversion as a least squares solver, which can be deteriorate accuracy as the size of the Volterra filter increases [33].

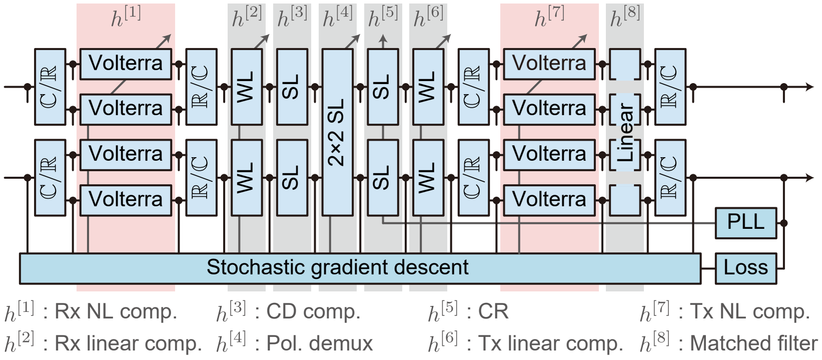


Figure 5.1. Architecture of adaptive ML filters including two Volterra layers to compensate for nonlinearity that occurs in Rx and Tx. Coefficients are adaptively controlled with back propagation and SGD to minimize loss. $\mathbb{C}/\mathbb{R}$: complex-to-real conversion, $\mathbb{R}/\mathbb{C}$: real-to-complex conversion.

5.2.1 Forward and back propagation of Volterra filter

We already have the forward and back propagation in the cases where the l-th layer in the adaptive multi-layer filter architecture shown in Fig. 5.1 consists of SL (MIMO) and WL (MIMO) filters. The remaining to obtain the last outputs of the multi-layer filters and the gradient calculation is the forward and back propagation in the case where the l-th layer has four real-valued Volterra filters for the IQ components.

Forward propagation

The output and input signal vectors of the l-th layer that relate to the last $L = 8$ layer outputs at the timing integer k are represented as $\boldsymbol{u}_i^{[l]}[k]$ and $\boldsymbol{u}_i^{[l-1]}[k]$, respectively, in the complex-valued phasor representation, as

$$\boldsymbol{u}_i^{[l]}[k] = (u_i^{[l]}[k], u_i^{[l]}[k-1], \ldots, u_i^{[l]}[k-M_l+1])^{\mathrm{T}}, \tag{5.1}$$

$$\boldsymbol{u}_i^{[l-1]}[k] = (u_i^{[l-1]}[k], u_i^{[l-1]}[k-1], \ldots, u_i^{[l-1]}[k-M_{l-1}+1])^{\mathrm{T}}, \tag{5.2}$$

where $i = 1, 2$ for the two polarizations, and M_l and M_{l-1} are the lengths of the output and input signal vectors, respectively. Given that the tap length of the FIR filters of the l-th layer is $M^{[l]}$, it satisfies

$$M^{[l]} = M_{l-1} - M_l + 1, \tag{5.3}$$

due to the relation of convolution, and it can also holds to Volterra filters. The l-th layer inputs are the $(l-1)$-th layer outputs in the multi-layer filter architecture. The last L-th layer outputs are $\boldsymbol{u}_i^{[L]}[k] = u_i^{[L]}[k]$ and $M_L = 1$. We obtain them by applying the forward propagation relation of each layer from the first $l = 1$ layer successively.

First, we require the relation between the signals in the complex-valued phasor representation and those in the real-valued IQ basis representation. The outputs of the l-th layer in the phasor representation $u_i^{[l]} \in \mathbb{C}$ are related to those in the IQ basis representation $u_{i\mathrm{I}}^{[l]}, u_{i\mathrm{Q}}^{[l]} \in \mathbb{R}$, by using the augmented signals with complex conjugates of $\underline{\boldsymbol{u}}^{[l]} = (u_1^{[l]}, u_2^{[l]}, u_1^{[l]*}, u_2^{[l]*})^{\mathrm{T}}$, as [42]

$$\begin{pmatrix} u_1^{[l]} \\ u_2^{[l]} \\ u_1^{[l]*} \\ u_2^{[l]*} \end{pmatrix} = \begin{pmatrix} 1 & 0 & i & 0 \\ 0 & 1 & 0 & i \\ 1 & 0 & -i & 0 \\ 0 & 1 & 0 & -i \end{pmatrix} \begin{pmatrix} u_{1\mathrm{I}}^{[l]} \\ u_{2\mathrm{I}}^{[l]} \\ u_{1\mathrm{Q}}^{[l]} \\ u_{2\mathrm{Q}}^{[l]} \end{pmatrix} = T_2 \begin{pmatrix} u_{1\mathrm{I}}^{[l]} \\ u_{2\mathrm{I}}^{[l]} \\ u_{1\mathrm{Q}}^{[l]} \\ u_{2\mathrm{Q}}^{[l]} \end{pmatrix}, \tag{5.4}$$

where

$$T_2 = \begin{pmatrix} I_2 & iI_2 \\ I_2 & -iI_2 \end{pmatrix} \tag{5.5}$$

satisfies $T_2^{\dagger} T_2 = T_2 T_2^{\dagger} = 2I_4$. I_n is an identity matrix with a corresponding size. We denote the signals in the IQ basis representation as $(u_{1\mathrm{I}}^{[l]}, u_{2\mathrm{I}}^{[l]}, u_{1\mathrm{Q}}^{[l]}, u_{2\mathrm{Q}}^{[l]})^{\mathrm{T}} = (x_1^{[l]}, x_2^{[l]}, x_3^{[l]}, x_4^{[l]})^{\mathrm{T}} = \boldsymbol{x}^{[l]}$. $x_q^{[l]} \in \mathbb{R}$, where $q = 1, \ldots, 4$, are corresponding IQ components of the two polarizations. According to Eq. (5.4),

$$\underline{\boldsymbol{u}}^{[l]} = T_2 \boldsymbol{x}^{[l]}, \tag{5.6}$$

and

$$\boldsymbol{x}^{[l]} = \frac{1}{2} T_2^{\dagger} \underline{\boldsymbol{u}}^{[l]}. \tag{5.7}$$

These relations hold for inputs.

When the l-th layer has four real-valued Volterra filters corresponding to $q = 1, \ldots, 4$, the output and input signal vectors are

$$\boldsymbol{x}_q^{[l]}[k] = (x_q^{[l]}[k], x_q^{[l]}[k-1], \ldots, x_q^{[l]}[k-M_l+1])^{\mathrm{T}}, \tag{5.8}$$

$$\boldsymbol{x}_q^{[l-1]}[k] = (x_q^{[l-1]}[k], x_q^{[l-1]}[k-1], \ldots, x_q^{[l-1]}[k-M_{l-1}+1])^{\mathrm{T}}. \tag{5.9}$$

If a Volterra filter includes up to the P-th order terms, the output is

$$x_q^{[l]}[k] = \sum_{p=1}^{P} \sum_{m_1=0}^{M^{[l]}-1} \sum_{m_2=m_1}^{M^{[l]}-1} \cdots \sum_{m_p=m_{p-1}}^{M^{[l]}-1} h_{q,p}^{[l]}[m_1, \ldots, m_p] \prod_{r=1}^{p} x_q^{[l-1]}[k-m_r], \tag{5.10}$$

where $h_{q,p}^{[l]}[m_1, \ldots, m_p] \in \mathbb{R}$ are the coefficients of the p-th order. Reflecting the symmetry of the product of $x_q^{[l-1]}$, a Volterra filter has coefficients of the number of $\sum_{p=1}^{P} \sum_{m_1=0}^{M^{[l]}-1} \sum_{m_2=m_1}^{M^{[l]}-1} \cdots \sum_{m_p=m_{p-1}}^{M^{[l]}-1} 1$. We can carry out the forward propagation in the cases where the l-th layer has Volterra filters by using Eq. (5.10).

Back propagation

By using Wirtinger derivatives, the complex-valued coefficients ξ of SL or WL filters are updated with gradient descent to minimize ϕ as

$$\xi \to \xi - 2\alpha \frac{\partial \phi}{\partial \xi^*}, \tag{5.11}$$

where α is a step size. When the l-th layer has real-valued Volterra filters, the update of a real-valued coefficient ξ with SGD is

$$\xi \to \xi - \alpha \frac{\partial \phi}{\partial \xi}, \tag{5.12}$$

instead of Eq. (5.11).

We require the relation between gradients in terms of the complex-valued phasor representation and those of the real-valued IQ basis representation as with the case of the forward propagation. According to Wirtinger derivatives,

$$\begin{pmatrix} \frac{\partial \phi}{\partial u_{1\mathrm{I}}^{[l]}} \\ \frac{\partial \phi}{\partial u_{2\mathrm{I}}^{[l]}} \\ \frac{\partial \phi}{\partial u_{1\mathrm{Q}}^{[l]}} \\ \frac{\partial \phi}{\partial u_{2\mathrm{Q}}^{[l]}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ i & 0 & -i & 0 \\ 0 & i & 0 & -i \end{pmatrix} \begin{pmatrix} \frac{\partial \phi}{\partial u_1^{[l]}} \\ \frac{\partial \phi}{\partial u_2^{[l]}} \\ \frac{\partial \phi}{\partial u_1^{[l]*}} \\ \frac{\partial \phi}{\partial u_2^{[l]*}} \end{pmatrix}, \tag{5.13}$$

and thus

$$\frac{\partial \phi}{\partial \boldsymbol{x}^{[l]}} = T_2^{\mathrm{T}} \frac{\partial \phi}{\partial \underline{\boldsymbol{u}}^{[l]}}, \tag{5.14}$$

$$\frac{\partial \phi}{\partial \underline{\boldsymbol{u}}^{[l]}} = \frac{1}{2} T_2^{*} \frac{\partial \phi}{\partial \boldsymbol{x}^{[l]}}. \tag{5.15}$$

This gives us the gradients of the loss in terms of the real-valued outputs of a Volterra filter. These relations also hold for the gradients in terms of the inputs.

When the gradients of the loss ϕ in terms of the output $\boldsymbol{x}_q^{[l]}[k]$ of a Volterra filter, represented as

$$\frac{\partial \phi}{\partial \boldsymbol{x}_q^{[l]}[k]} = \left(\frac{\partial \phi}{\partial x_q^{[l]}[k]}, \frac{\partial \phi}{\partial x_q^{[l]}[k-1]}, \ldots, \frac{\partial \phi}{\partial x_q^{[l]}[k-M_l+1]} \right)^{\mathrm{T}}, \tag{5.16}$$

are given, the gradients in terms of the coefficients of the Volterra filter are

$$\frac{\partial \phi}{\partial h_{q,p}^{[l]}[m_1,\dots,m_p]} = \sum_{m_r=0}^{M_l-1} \frac{\partial \phi}{\partial x_q^{[l]}[k-m_r]} \frac{\partial x_q^{[l]}[k-m_r]}{\partial h_{q,p}^{[l]}[m_1,\dots,m_p]}, \tag{5.17}$$

$$\frac{\partial x_q^{[l]}[k-m_r]}{\partial h_{q,p}^{[l]}[m_1,\dots,m_p]} = x_q^{[l-1]}[k-m_r-m_1]\cdots x_q^{[l-1]}[k-m_r-m_p], \tag{5.18}$$

and the gradients in terms of the input of the Volterra filter are

$$\frac{\partial \phi}{\partial x_q^{[l-1]}[k-m_s]} = \sum_{m_r=0}^{M_l-1} \frac{\partial \phi}{\partial x_q^{[l]}[k-m_r]} \frac{\partial x_q^{[l]}[k-m_r]}{\partial x_q^{[l-1]}[k-m_s]}, \tag{5.19}$$

$$\begin{aligned}
\frac{\partial x_q^{[l]}[k-m_r]}{\partial x_q^{[l-1]}[k-m_s]} =& \sum_{p=1}^{P} \Bigg(\sum_{m_2=m_s-m_r}^{M^{[l]}-1} \sum_{m_3=m_2}^{M^{[l]}-1} \cdots \sum_{m_p=m_{p-1}}^{M^{[l]}-1} h_{q,p}^{[l]}[m_s-m_r, m_2, \dots, m_p] \\
& \cdot x_q^{[l-1]}[k-m_r-m_2]\cdots x_q^{[l-1]}[k-m_r-m_p] \\
& + \sum_{m_1=0}^{m_s-m_r} \sum_{m_3=m_s-m_r}^{M^{[l]}-1} \sum_{m_4=m_3}^{M^{[l]}-1} \cdots \sum_{m_p=m_{p-1}}^{M^{[l]}-1} h_{q,p}^{[l]}[m_1, m_s-m_r, m_3, \dots, m_p] \\
& \cdot x_q^{[l-1]}[k-m_r-m_1] x_q^{[l-1]}[k-m_r-m_3]\cdots x_q^{[l-1]}[k-m_r-m_p] \\
& + \cdots \\
& + \sum_{m_1=0}^{m_s-m_r} \sum_{m_2=m_1}^{m_s-m_r} \cdots \sum_{m_{p-1}=m_{p-2}}^{m_s-m_r} h_{q,p}^{[l]}[m_1, \dots, m_{p-1}, m_s-m_r] \\
& \cdot x_q^{[l-1]}[k-m_r-m_1]\cdots x_q^{[l-1]}[k-m_r-m_{p-1}] \Bigg).
\end{aligned} \tag{5.20}$$

Now we obtain the back propagation of the real-valued Volterra filters. This includes the case of the linear filter (as is in the case of the eighth layer for matched filtering in Fig. 5.1) if we restrict $P = 1$. Therefore, we now have all the forward and the back propagation of the adaptive multi-layer filters including the Volterra filter layers.

5.3 Evaluation in simulation

We evaluated the performance of compensation of Tx and Rx nonlinearity by the adaptive multi-layer filters including Volterra filter layers shown in Fig. 5.1 through simple numerical simulations.

Simulation model

The simulation model corresponds to a transmission system with coherent detection as shown in Fig. 2.2, though we consider a case without WDM to focus on Tx and Rx nonlinearity. The transmitted signal was a 23 Gbaud[3] PDM-64QAM.

---

[3] This symbol rate was chosen to mitigate the impact of linear Tx and Rx impairments remaining in the experimental setup described later to focus on the Tx and Rx nonlinearity.

Three FEC frames of low-density parity-check code for DVB-S2 with a frame length of 64,800 and code rate of 0.8 with loading random bits to its payload were generated for each IQ component of the two polarizations, and they were mapped into the PDM-64QAM. To conduct a coefficient update with pilot-based data-aided LMS, a pilot symbol of the 64QAM was inserted in every 15 symbols. In addition, a preamble of 64 successive symbols was inserted in each FEC frame to detect the head of the frame to carry out pilot-based signal processing. The preamble was quadrature phase-shift keying (QPSK) corresponding to the corner states of the 64QAM to distinguish it in a simple manner by using power difference. This results in about 6.5% overhead in total in terms of the pilot and preamble symbols. Learning of large nonlinear filters can fall into overfitting, especially when using a pseudo random sequence [105, 106]. In this study, a long random pattern was used for pilot and preamble symbols compared to the tap length of the Volterra filters and the performance was evaluated by a sufficient long random data load. The PDM-64QAM signal inserted with the pilot and preamble symbols was then upsampled to 32-fold oversampling, and a root raised cosine filter with a roll-off of 0.1 was performed. Real-valued IQ components of the two polarizations of this signal correspond to the electrical signals generated by the ideal DAC. An optical modulator driven by these signals modulated a CW light source at a frequency of 193.3 THz and linewidth of 0 to generate optical signals. At this time, Tx nonlinearity that was sinusoidal, i.e. $\sin(\beta x)/\beta$ with the degree of nonlinearity β, which emulated a characteristic of a Mach-Zehnder modulator, was imposed to the IQ components. No IQ linear impairments were imposed in the Tx to focus on the nonlinearity.

After adding CD corresponding to the 100-km SMF and setting an OSNR to 30 dB/0.1 nm by adding an additive white Gaussian noise, the signal was coherently received. A LO first had a linewidth of zero and no frequency offset to the signal carrier so that only CD was mutually non-commutative linear impairment with Tx or Rx nonlinearity [4]. A low-pass filter with a 3-dB bandwidth of 0.8 times of the symbol rate was used to filter the electrical signals after coherent detection. Rx nonlinearity, which is like a saturation characteristic of an electronic amplifier, i.e. $\tanh(\beta x)/\beta$, was imposed to the IQ components. No IQ linear impairments were imposed in the Rx again. After that, the signals were sampled with two-fold oversampling. Then, DSP was performed after normalizing IQ components individually.

In this simulation, two kind of architectures were evaluated. The first one is a conventional one used in optical fiber communication systems with coherent detection as a reference. This is referred to as Linear after only linear filters were used for impairment compensation[5]. In this case, the received signals were resampled to two-fold oversampling on the basis of a timing error [86] with low-pass filtering that uses 3-dB bandwidth of the symbol rate (which was not necessary in this simulation and worked in the experiment described later). Then, CD compensation and matched filtering were performed in the frequency domain. The head of the preamble symbols was detected by the moving average of the signal power, and the signals were aligned to perform pilot-based data-aided LMS for polarization demultiplexing and PMD compensation with a half-symbol spaced 21-tap FIR 2×2 MIMO filter. Carrier recovery was also

---

[4] The simulation results when a LO having a linewidth of 100 kHz was also included for reference.

[5] This configuration is the same as 2×2 SL in the previous chapters.

performed by a digital PLL within the MIMO filter.

The second architecture was the adaptive multi-layer filters including Volterra filter layers. Since CD compensation and matched filtering were performed in the multi-layer filter in this case, the timing alignment was performed after resampling. Here, the head of the preamble symbols was detected after CD compensation and the timing alignment was performed on the signal before CD compensation. Impairment compensation by the multi-layer filters was executed on the signal before CD compensation with the timing alignment again. The tap lengths of the second and sixth layers to compensate for linear impairment in the Rx and Tx were set to five. The length of the third layer to compensate for CD was 41, which was sufficient to compensate for the CD amount over a 100-km SMF. The length of the fourth layer to perform polarization demultiplexing and PMD compensation was 21. The length of the eighth layer was nine, though it was a bit insufficient to represent a root raised cosine filter with a roll-off of 0.1 precisely, since the increase of the tap length of later layers in the multi-layer filters results in high computational complexity. The lengths of the first and seventh layers of Volterra filters to compensate for Rx and Tx nonlinearity were one in this simulation, considering that the imposed nonlinearity in the Tx and Rx does not have any temporal spread in this simulation. The Volterra filters had the terms up to the fifth order. The coefficients of the first, second, fourth, sixth, and seventh layers were adaptively controlled to minimize the loss of the data-aided LMS with back propagation and SGD. They were updated at both the pilot and preamble symbols. The coefficients were initialized as one at the center of the main diagonal filters with the remaining set to zero. The fifth layer for carrier recovery was updated by a PLL at all the symbols. A PLL with a data-aided or decision-directed phase error with ideal one symbol delay was used to focus on the performance of adaptive control of the filter coefficients with SGD. A pilot-based digital PLL can be implemented with low penalty [107] and sophisticated parallelization techniques have been investigated to reduce loop delay [108], though further consideration is needed for practical implementation of a digital PLL. The step sizes were 10^{-3} for the second, fourth, and sixth layers, and 10^{-2} for the first and seventh Volterra filter layers. The same step size was used for all orders of Volterra filter coefficients. We evaluated three cases in the adaptive ML filter architecture; only the first layer to compensate for Rx nonlinearity was activated (ML Rx Volterra); only the seventh layer to compensate for Tx nonlinearity was activated (ML Tx Volterra); and both the first and seventh layers were activated (ML both Volterra). The four cases, which includes Linear, were compared.

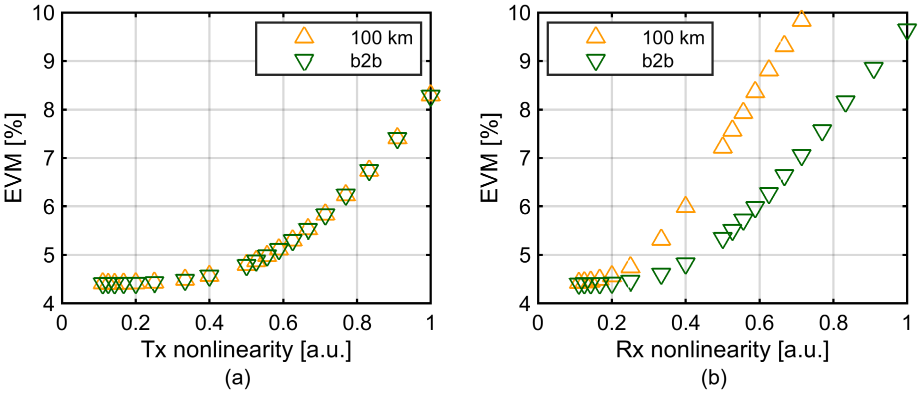
After impairment compensation and removing pilot and preamble symbols, the EVM was evaluated in this simulation. The error vector was defined as the difference between a signal to a decision result, and the reference of the EVM was the averaged amplitude of QAM.

Figure 5.2 shows the results of the EVM after the Linear processing while changing the degree of nonlinearity β in the Tx and Rx. Two cases with a back-to-back condition (b2b) and a CD accumulation of 100-km SMF (100 km) were evaluated. In the case with Tx nonlinearity, no difference was observed for the b2b and 100 km conditions. In contrast, in the case with Rx nonlinearity, the EVM got worse when it imposed in the 100 km condition[6]. This is because

---

[6] The difference between the cases of Tx and Rx nonlinearity is the consequence of the

Rx nonlinearity more severely affects for a signal with a large amplitude distribution, i.e., a large peak-to-average ratio (PAPR), after CD accumulation than for one with a small PAPR before CD accumulation. This phenomenon does not occur for Tx nonlinearity as we can see in the order of impairments. Tx and Rx nonlinearity differently when other impairments exist. In the following evaluations, the degree of nonlinearity was set to $\beta = 0.83$ for Tx nonlinearity and to 0.67 for Rx nonlinearity to provide similar EVM degradation in the b2b condition.



(a)

(b)

Figure 5.2. Emulation of (a) Tx and (b) Rx nonlinearity in the cases with back-to-back (b2b) and 100-km SMF transmission (100 km). The results of EVM averaged over two polarizations after impairment compensation by the Linear processing are plotted.

Back-to-back condition

We evaluated the performance of the four compensation cases (Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra) under four conditions: without nonlinearity in the Tx and Rx (no NL), with Tx nonlinearity (Tx NL), with Rx nonlinearity (Rx NL), and with both Tx and Rx nonlinearity (both NL). Figure 5.3 shows the constellation results in the b2b condition obtained by the four compensation cases. Only the constellations of one polarization are shown in Fig. 5.3 for simplicity, while a similar result was obtained for the orthogonal polarization. The results by Linear processing show that the constellation of 64QAM was distorted, where spacing between the symbol points became narrower at the outer symbol points, when Tx nonlinearity or Rx nonlinearity was imposed. In the case with both Tx and Rx nonlinearity, the constellation was distorted more severely in the same way. In contrast, the results by ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing provided constellations of almost equally-spaced symbol points, as was the case without nonlinearity, when Tx and Rx nonlinearity were imposed. This indicates that three kinds of ML Volterra processing could compensate for both Tx and Rx

---

difference of the nonlinear function we adopted.

nonlinearity, though nonlinear compensation did not perfectly work to cancel Tx nonlinearity. In other words, each of the first and seventh Volterra filter layers of the ML filters could compensate for both Tx and Rx nonlinearity in the b2b condition. It should be noted that this was achieved only if no other effects, such as CD, polarization rotation, phase noise from a LO, and a frequency offset, coexist as shown later.

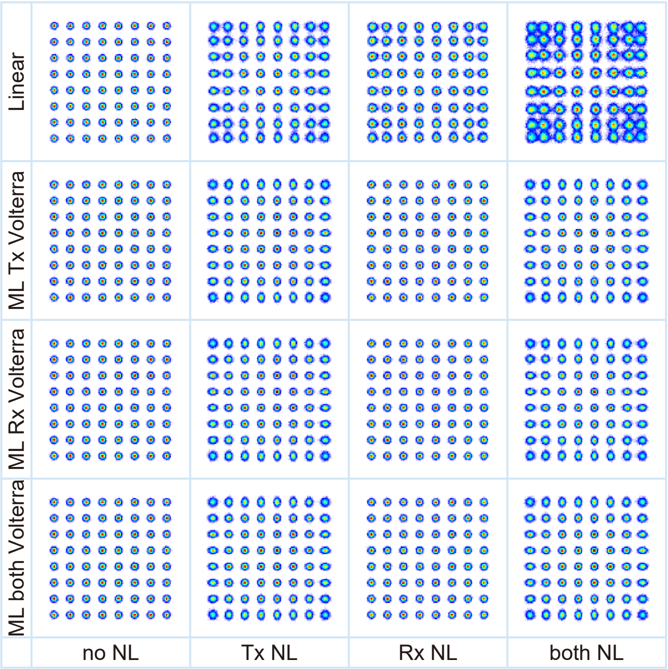


Figure 5.3. Simulation results of constellations in the b2b condition obtained by Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing, in the cases with no nonlinearity (no NL), Tx nonlinearity (Tx NL), Rx nonlinearity (Rx NL), and both Tx and Rx nonlinearity (both NL).

100-km SMF transmission

Figure 5.4 shows the constellation results obtained by the four compensation cases in a 100-km transmission. The results by Linear processing show that the constellation under Tx nonlinearity was similar to that in the b2b condition shown in Fig. 5.3. Whereas, the constellation under Rx nonlinearity was not similar to that in the b2b condition, and it had equally-spaced symbol points but with increased errors around the symbol points. This is because Rx nonlinearity was imposed under CD accumulation and compensation for CD was performed after that. In the constellation with both Tx and Rx nonlinearity, both effects of distortion of unequally-spaced symbol points and increased errors around symbol points were superimposed. In the case by ML Tx Volterra processing, it compensated for Tx nonlinearity, whereas it did not compensate for Rx nonlinearity. In the case by ML Rx Volterra processing, it compensated for Rx nonlinearity, whereas it did not compensate for Tx nonlinearity. The results showed that Tx and Rx nonlinearity differed in their behaviors and that compensation of Tx or Rx nonlinearity could not work well unless the nonlinear compensation was positioned in an appropriate order in impairment compensation when other effects, which is CD in this case, coexist due to mutual non-commutativity of the effects. Since two nonlinear compensations of Rx and Tx nonlinearity were included in the case by ML both Volterra processing, it compensated for both Tx and Rx nonlinearity as shown in Fig. 5.4.

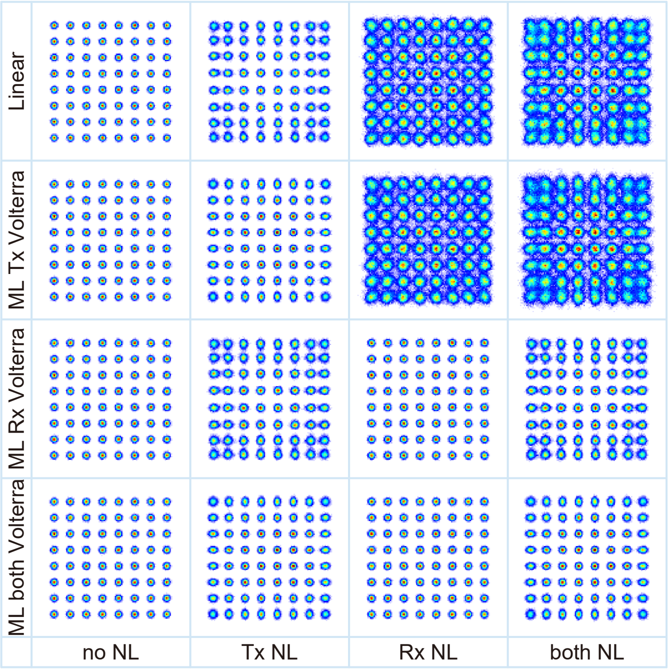


Figure 5.4. Simulation results of constellations in 100-km SMF transmission obtained by Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing, in the cases with no nonlinearity (no NL), Tx nonlinearity (Tx NL), Rx nonlinearity (Rx NL), and both Tx and Rx nonlinearity (both NL).

Comparison of various conditions

Figure 5.5 summarizes the EVM results of the constellations shown in Figs. 5.3 and 5.4. In the b2b condition shown in Fig. 5.5(a), the EVMs obtained by four compensation cases with no Tx or Rx nonlinearity were similar and it indicated that a small number of taps for the eighth layer provided a small penalty in the cases with multi-layer filters. The EVMs obtained by ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing were similar and improved from those by Linear processing when Tx and Rx nonlinearity were imposed. In the case with both Tx and Rx nonlinearity, the EVM by Linear processing was 9.3% and that by ML both Volterra was 5.4%. The performance by ML both Volterra did not reach the performance without nonlinearity by Linear processing of 4.4%. Possible causes of this are the limitations of nonlinear processing with two-fold oversampling, low-pass filtering at the Rx, and restrictions of Volterra filters up to the fifth order. In the 100-km transmission shown in Fig. 5.5(b), ML Tx Volterra and ML both Volterra processing provided good performances of 5.3% under Tx nonlinearity, and ML Rx Volterra and ML both Volterra processing provided good ones of 4.7% under Rx nonlinearity. The EVM by ML both Volterra processing under both Tx and Rx nonlinearity was 5.5%. These results demonstrated that the proposed adaptive ML filters including Volterra filter layers could work for compensating for both Tx and Rx nonlinearity simultaneously and effectively even when other impairments coexist.

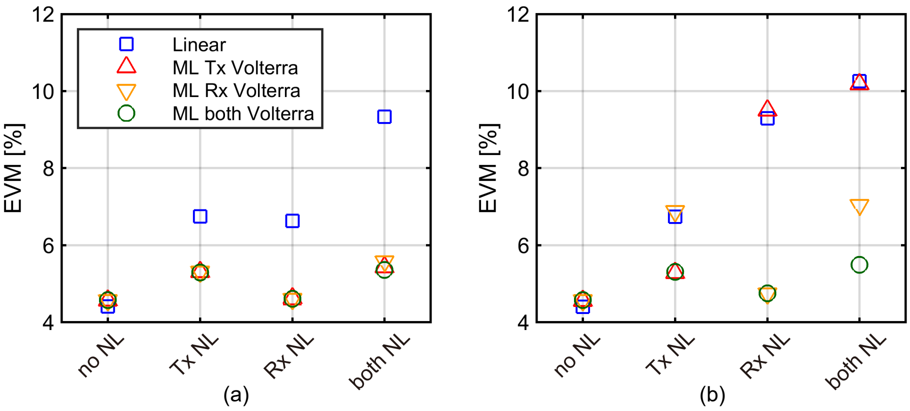


Figure 5.5. Simulation results of EVM obtained by Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing, where no nonlinearity (no NL), Tx nonlinearity (Tx NL), Rx nonlinearity (Rx NL), and both Tx and Rx nonlinearity (both NL) were imposed, in the cases of (a) b2b and (b) 100-km transmission.

In the previous simulations, a laser phase noise was ignored so that only CD was mutually non-commutative linear impairment with Tx or Rx nonlinearity. There were no mutually non-commutative linear impairments with Tx or Rx nonlinearity in the b2b condition, in comparison to the case when CD coexisted. Figure 5.6 shows the simulation results of EVM when an LO linewidth was 100 kHz. In contrast to the case without a phase noise shown in Fig. 5.5(a), the EVMs obtained by ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing were different in the b2b condition as shown in Fig. 5.6(a) when Tx or Rx nonlinearity was imposed. Tx nonlinearity was compensated by ML Tx Volterra and ML both Volterra processing. Rx nonlinearity was compensated by ML Rx Volterra and ML both Volterra processing. The EVM results in the 100-km transmission shown in Fig. 5.6(b) show similar behaviors. Thus, not only CD but also other linear effects that are not commutative with Tx or Rx nonlinearity prevent one nonlinear filter positioned at the first or the last of compensation blocks from compensating for both Tx and Rx nonlinearity. Nevertheless, ML both Volterra processing could compensate for all the relevant impairments including Tx and Rx nonlinearity.

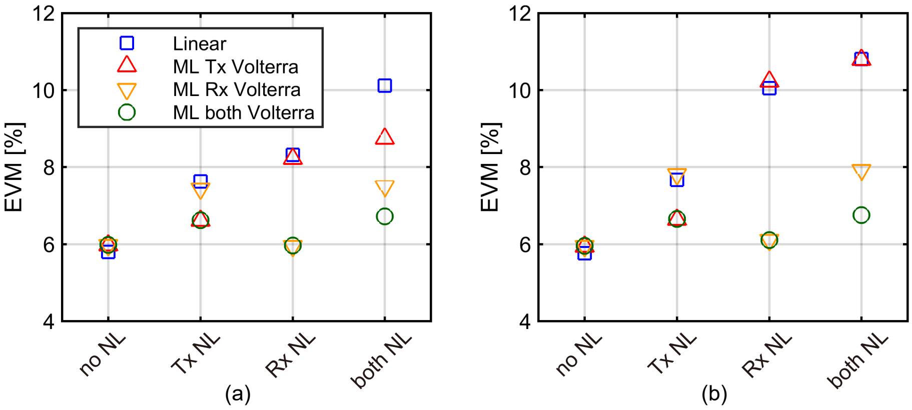


Figure 5.6. Simulation results of EVM with LO phase noise of 100 kHz obtained by Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing, where no nonlinearity (no NL), Tx nonlinearity (Tx NL), Rx nonlinearity (Rx NL), and both Tx and Rx nonlinearity (both NL) were imposed, in the cases of (a) b2b and (b) 100-km transmission.

Convergence of loss

We also evaluated the convergence property of the adaptive coefficient control. Figure 5.7 shows the time development of a training loss, that is, the loss at pilot and preamble symbols, in the 100-km transmission. In all cases of Tx and Rx nonlinearity, convergence was achieved by updating with about 2×10^4 symbols. There were multiple peaks observed in the loss when Tx or Rx nonlinearity was imposed. These peaks were caused by preamble symbols. The four corner states of the 64QAM were used in preamble symbols for timing alignment. Preamble symbols had higher averaged power compared to the other symbols and thus induced more severe degradation when nonlinearity was imposed.

5.4 Evaluation in transmission experiment

Finally, we evaluated the performance of the compensation of Tx and Rx nonlinearity by the adaptive multi-layer filters including Volterra filter layers in a transmission experiment of a 23 Gbaud PDM-64QAM signal over a 100-km SMF span. In this experiment, Tx and Rx nonlinearity were induced by tuning an output amplitude of electronic amplifiers in a Tx and Rx.

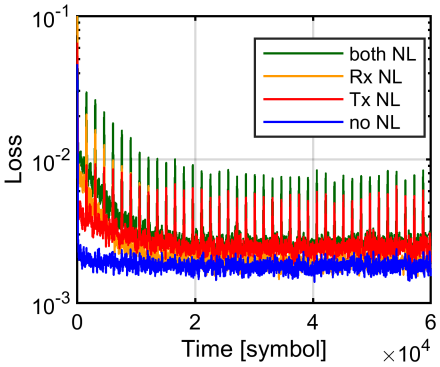


Figure 5.7. Simulation results of time development of training loss by ML both Volterra processing moving-averaged over 100 symbols.

Experimental setup

The experimental setup is shown in Fig. 5.8. At the transmitter side, as was the case with the previous simulation, three FEC frames were generated for each IQ component of the two polarizations, and they were mapped into the PDM-64QAM with the inserted pilot and preamble symbols. Real-valued IQ components of the two polarizations of the signal were then upsampled to two-fold oversampling, and a raised cosine filter with a roll-off of 0.1 and pre-compensation of frequency characteristics of Tx components were performed. These signals were upsampled again to four-fold oversampling to be generated by a DAC at a sampling rate of 92 GS/s and a vertical resolution of eight bits. An optical modulator driven by the outputs of the DAC with an LD source at a frequency of 193.3 THz having a linewidth of about 100 kHz generated a 23 Gbaud PDM-QPSK signal. Tx nonlinearity was induced by tuning the output amplitude of electronic driver amplifiers operating on the DAC outputs.

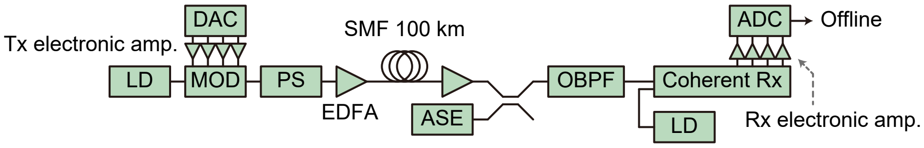


Figure 5.8. Experimental setup for transmission of a 23 Gbaud PDM-64QAM signal over the span of a 100-km SMF. Tx(Rx) nonlinearity was induced by tuning an output amplitude of electronic amplifiers in Tx(Rx).

The generated signal was then transmitted over a span of a 100-km SMF after low speed ($10 \times 2\pi$ rad/s) polarization scrambling. The span input power was set to 0 dBm, providing some fiber nonlinearity of self-phase modulation.

After the 100-km SMF transmission, the OSNR was set by adding ASE. In the case when no ASE was added, the received OSNR was 34.9 dB/0.1 nm. The signal then passed through an optical band-pass filter with a 3-dB bandwidth of 50 GHz and was detected with an integrated polarization-diversity coherent Rx including gain controlled electronic amplifiers. Its 3-dB bandwidth was 40 GHz. An LO with a linewidth of about 100 kHz and the signal source were free-running, and the frequency offset was within about ± 100 MHz. Since a frequency offset can be estimated and compensated after CD compensation [72], this small frequency offset is a reasonable start point of the adaptive control of filters. The electrical signals of the coherent Rx outputs were sampled by a digital oscilloscope as an ADC at a sampling rate of 80 GS/s, vertical resolution of eight bits, and bandwidth of 25 GHz. Three acquisitions were obtained for each condition. Rx nonlinearity was induced by tuning the output amplitude of the TIA in the coherent Rx.

DSP was performed offline. In this experiment, we also evaluated the performance of four compensation cases (Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra). The configuration of DSP was the same as that in the previous simulation, except that the tap length of the fourth layer to perform polarization demultiplexing and PMD compensation (and also that in Linear) was 41. The tap lengths and the maximum orders of Volterra filters of the first and seventh layers to compensate for Rx and Tx nonlinearity were optimized since the temporal spread of nonlinearity was unknown in this experiment, as discussed later. The step sizes including the Volterra filter layers were 10^{-3}. After convergence of the adaptive control of the filter coefficients by using about 3×10^6 samples, which includes about 6×10^4 pilot and preamble symbols, the pre-/post-FEC bit error rate (BER) and EVM were evaluated with the removal of the pilot and preamble symbols. The median of the results of the three acquisitions was adopted.

Figure 5.9 shows the experimental results of the EVM obtained by Linear processing when the output amplitude of electronic driver amplifiers in the Tx or that of the TIA in the Rx was tuned. The output amplitude was normalized by its optimum value. In the cases when the electronic output amplitudes were tuned in both the Tx and Rx, the EVM degraded as the output amplitude exceeded its optimum, which suggested that some nonlinearity was induced in the Tx and Rx. The optical spectra when the output amplitude of the Tx electronic amplifiers was tuned are shown in Fig. 5.10. The powers of the four cases were normalized. As the output amplitude of the Tx electronic amplifiers increased, the optical spectrum became slightly broader due to nonlinearity. We hereinafter evaluated the performance in the four conditions similar to that in the simulation: the output amplitudes were both optimum and no nonlinearity was induced in the Tx and Rx (no NL); the output amplitude in the Tx was two and Tx nonlinearity was induced (Tx NL); the output amplitude in the Rx was two and Rx nonlinearity was induced (Rx NL); the output amplitudes in the Tx and Rx were two and both Tx and Rx nonlinearity were induced (both NL).

Dependence of performance on order and length of Volterra filters

We evaluated the EVM while changing the tap lengths and maximum orders of the Volterra filters to optimize them. No ASE was added after transmission.

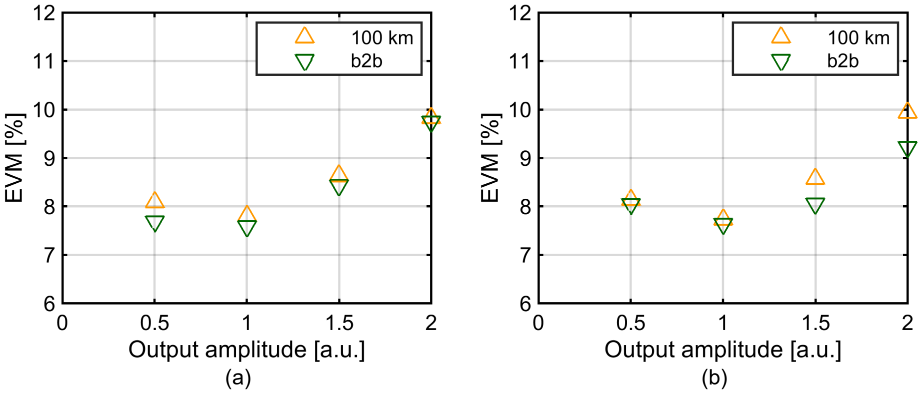


Figure 5.9. Experimental results of EVM obtained by Linear processing when the output amplitude of the (a) Tx or (b) Rx electronic amplifiers was tuned. The output amplitude was normalized by its optimum value.

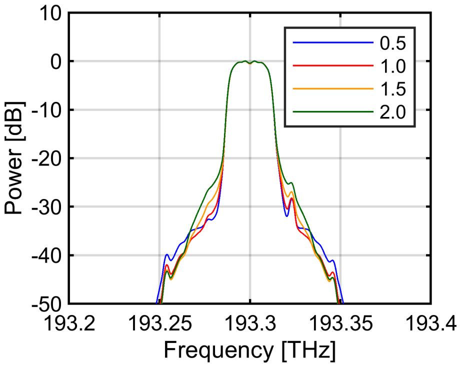
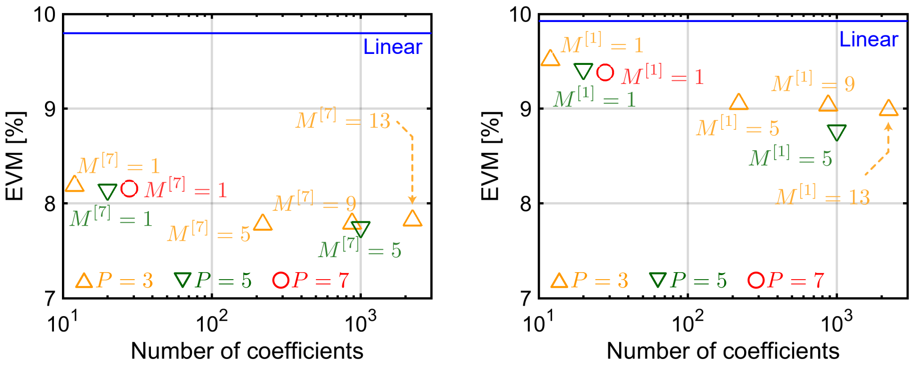


Figure 5.10. Optical spectra when the output amplitude of the Tx electronic amplifiers was tuned from 0.5 to 2.0. The powers of the four cases were normalized. The resolution bandwidth was 0.02 nm.

Figure 5.11(a) shows the results of the EVM by ML Tx Volterra processing where Tx nonlinearity was induced, as a function of the total number of coefficients of the Volterra filters of the seventh layer. The tap length $M^{[7]}$ and the maximum order P of the seventh layer Volterra filters were changed. The step size was roughly optimized for each size. Focusing on the results with the cases where the tap length of the Volterra filters $M^{[7]} = 1$, the performance with $P = 5$ outperformed that with $P = 3$, and little improvement was observed with $P = 7$. Then, focusing on the results when the maximum order $P = 3$, they indicated that a tap length of five was a good choice. Figure 5.11(b) shows the results of EVM by ML Rx Volterra processing where Rx nonlinearity was induced. The tap length and maximum order of the first layer Volterra filters were changed. In accordance with these results, we set the tap lengths of the two Volterra filter layers to $M^{[7]} = M^{[1]} = 5$, and the maximum orders $P = 5$.



(a) (b)

Figure 5.11. Experimental results of dependence of EVM on Volterra filter size, (a) obtained by ML Tx Volterra processing where Tx nonlinearity was induced, and (b) by ML Rx Volterra processing where Rx nonlinearity was induced.

Figure 5.12 shows the histograms of the square amplitudes of converged coefficients of the four Volterra filters at the first and seventh layers obtained by ML both Volterra processing in the 100 km transmission experiment with Tx and Rx nonlinearity induced (both NL) when no ASE was added. The tap lengths of the two Volterra filter layers were $M^{[7]} = M^{[1]} = 5$, and the maximum orders were $P = 5$. It is noted that the inputs of the multi-layer filters were normalized so that the corner states of 64QAM corresponded to the square amplitude of one. The histograms of the coefficients in the order of $p = 1$ to 5 are individually shown. Clear discrimination to choose important coefficients could not be extracted regarding both the Volterra filters of the seventh layer for Tx nonlinearity compensation and the ones of the first layer for Rx nonlinearity compensation with the LMS algorithm. Regularization techniques such as L_1 regularization can induce sparse coefficients and pruning of them may help to reduce the computational complexity of the Volterra filters.

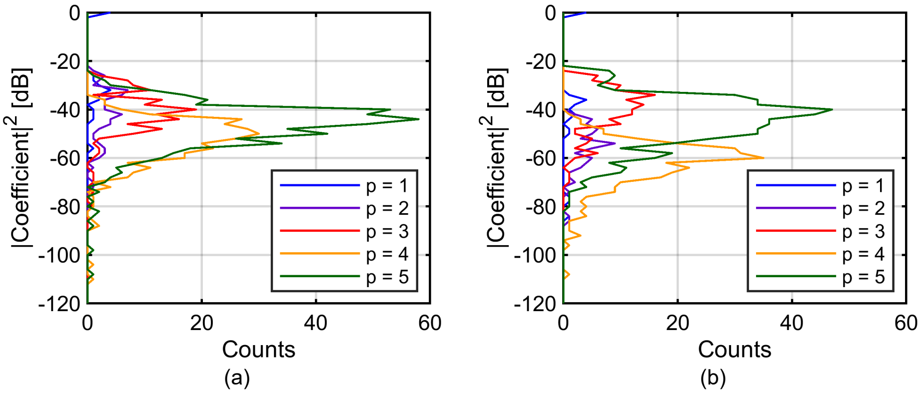


Figure 5.12. Histogram of the square amplitude of converged coefficients of the Volterra filters of (a) the seventh layer for Tx nonlinearity compensation and (b) the first layer for Rx nonlinearity compensation when ML both Volterra processing was used in the 100 km transmission experiment with Tx and Rx nonlinearity (both NL).

Tx and Rx nonlinearity compensation by adaptive multi-layer filters including Volterra filters

Figure 5.13 shows the results of constellations obtained after the 100-km transmission by the four compensation cases when Tx or Rx nonlinearity was induced. As was the case in the simulation, ML Tx Volterra processing compensated for Tx nonlinearity and ML Rx Volterra processing compensated for Rx nonlinearity. In the case of ML both Volterra processing, compensation of both Tx and Rx nonlinearity was accomplished. Moreover, ML Tx Volterra and ML both Volterra processing provided better constellations in the case without nonlinearity induced compared with that by Linear processing. This suggests that some Tx nonlinearity remained when the electronic output amplitude in the Tx was at its optimum in this experiment.

We evaluated the performance of the four compensation cases while changing the received OSNR to assess the robustness of the adaptive control of the multi-layer filter coefficients including Volterra filter layers with back propagation and SGD. Figures 5.14 and 5.15 show the results of pre-/post-FEC BER, respectively, obtained by the four compensation cases in the four conditions of Tx and Rx nonlinearity. The error-free results are plotted at 10^{-6} for visibility. In the condition without Tx and Rx nonlinearity induced (Figs. 5.14(a) and 5.15(a)), ML both Volterra and ML Tx Volterra processing provided better pre-FEC BER than the others, especially at a high OSNR region, since Tx nonlinearity remained in this case. The post-FEC error-free was achieved above 22 dB/0.1 nm with all four compensation cases. In the condition with Tx nonlinearity (Figs. 5.14(b) and 5.15(b)), the results of the pre-FEC BER showed similar behaviors in the case with no nonlinearity induced, except that they were all degraded. The post-FEC error-free was achieved down to 29 dB/0.1

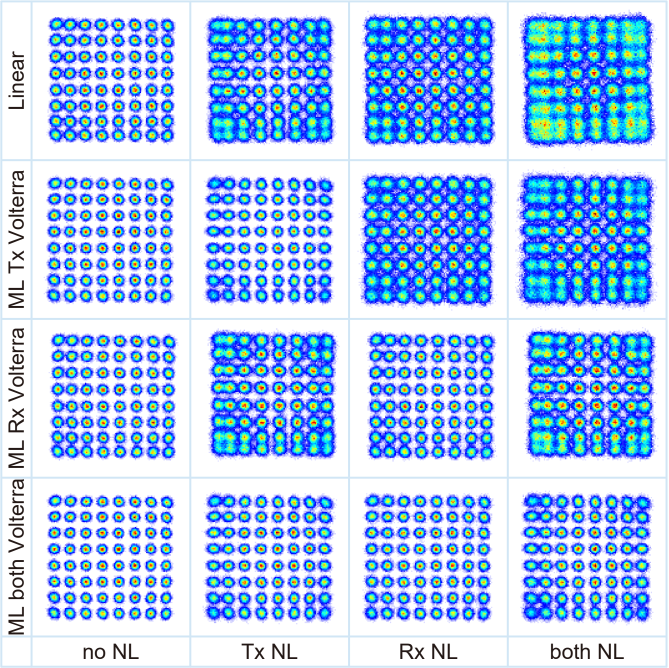


Figure 5.13. Experimental results of constellations in 100-km transmission obtained by Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing in the conditions with no nonlinearity induced (no NL), Tx nonlinearity induced (Tx NL), Rx nonlinearity induced (Rx NL), and both Tx and Rx nonlinearity induced (both NL).

nm by ML Rx Volterra processing, 25 dB/0.1 nm by Linear, and 23 dB/0.1 nm by ML Tx Volterra and ML both Volterra processing. When no nonlinearity or Tx nonlinearity was induced, the performance of ML Rx Volterra processing was slightly worse than that of Linear processing. The possible cause of this might be excess errors in the adaptive coefficient control [41]. Since ML Rx Volterra has more coefficients than Linear processing, the degradation due to excess errors was more severe. In the condition with Rx nonlinearity (Figs. 5.14(c) and 5.15(c)), ML both Volterra processing provided the best pre-FEC BER, followed by ML Rx Volterra processing since it worked to compensated for Rx nonlinearity. The post-FEC error-free was achieved down to 25 dB/0.1 nm by Linear and ML Tx Volterra, 24 dB/0.1 nm by ML Rx Volterra, and 22 dB/0.1 nm by ML both Volterra processing. In the condition with both Tx and Rx nonlinearity (Figs. 5.14(d) and 5.15(d)), the post-FEC error-free was not achieved by Linear processing, and it was achieved down to 30 dB/0.1 nm by ML Rx Volterra, 26 dB/0.1 nm by ML Tx Volterra, and 23 dB/0.1 nm by ML both Volterra processing. ML both Volterra processing could provide the better robust performance compared with the others regardless of Tx or Rx nonlinearity induced. Therefore, the results demonstrated that the adaptive ML filters including Volterra filter layers could compensate for the nonlinearity that occurs in both Tx and Rx simultaneously and effectively under the accumulation of CD and other effects including a frequency offset and polarization rotation, which are non-commutative impairments with Tx and Rx nonlinearity.

Robustness of learning of Volterra filters to OSNR condition

The qualities of the adaptive control of the Volterra filter coefficients in different OSNR conditions were assessed. Figure 5.16 shows the experimental results of the pre-/post-FEC BER in 100-km transmission obtained by ML both Volterra processing in the conditions with both Tx and Rx nonlinearity induced (both NL). We compared the performances in three adaptive control procedures when changing the OSNR; the coefficients of the Volterra filters at the first and seventh layers were converged at each corresponding received OSNR individually, which was the same as the case in obtaining the previous experimental results; the converged coefficients obtained at the highest OSNR (without ASE) were used for the Volterra filters and they were fixed without update at each OSNR; and the coefficients were initialized as the converged coefficients obtained at the highest OSNR and they were also updated at each OSNR. According to Fig. 5.16, these three cases provided similar pre- and post-BERs down to low OSNR. Thus, adaptive control of the Volterra filter coefficients was robust and gave little penalty at low OSNR in this evaluation.

5.5 Summary

We proposed an multi-layer filter architecture in which two Volterra filter layers were positioned to compensate for nonlinearity that occurs in Tx and Rx components under other impairments such as CD. The coefficients of the multi-layer filters including Volterra filter layers are adaptively controlled by using a gradient calculation with back propagation and SGD to minimize the loss function that is composed of the last layer outputs. We evaluated the performance of

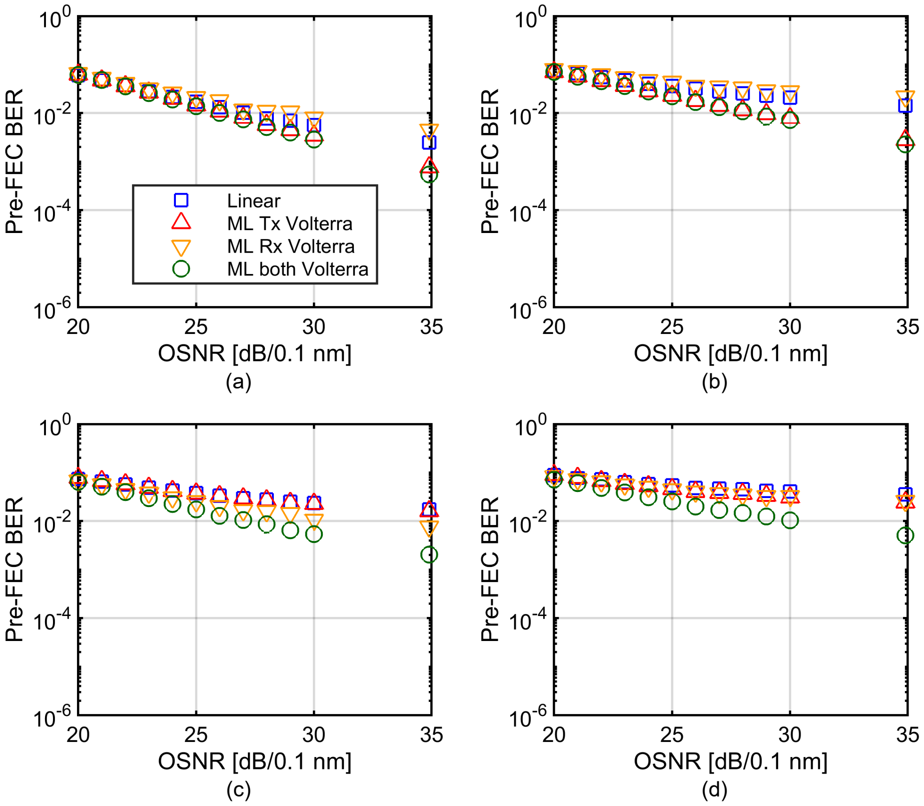


Figure 5.14. Experimental results of dependence of pre-FEC BER on OSNR obtained by Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing in the conditions with (a) no nonlinearity induced (no NL), (b) Tx nonlinearity induced (Tx NL), (c) Rx nonlinearity induced (Rx NL), and (d) both Tx and Rx nonlinearity induced (both NL).

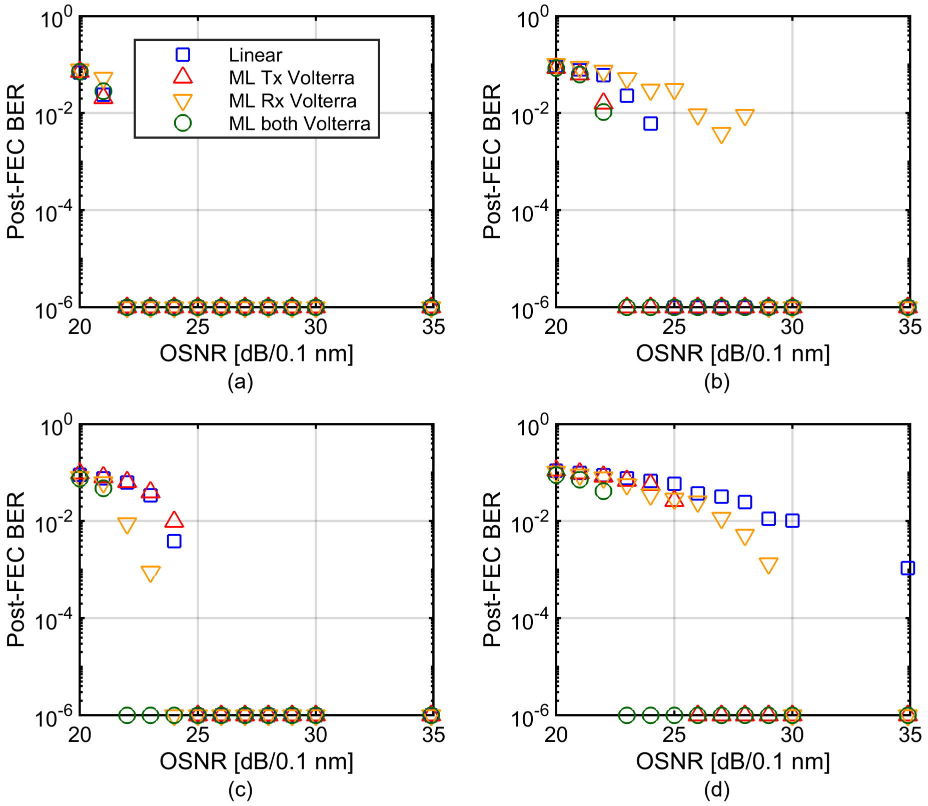


Figure 5.15. Experimental results of dependence of post-FEC BER on OSNR obtained by Linear, ML Tx Volterra, ML Rx Volterra, and ML both Volterra processing in the conditions with (a) no nonlinearity induced (no NL), (b) Tx nonlinearity induced (Tx NL), (c) Rx nonlinearity induced (Rx NL), and (d) both Tx and Rx nonlinearity induced (both NL).

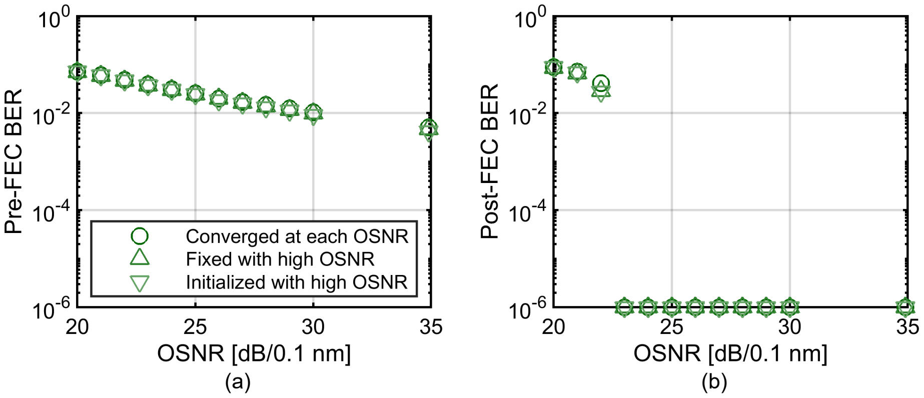


Figure 5.16. Experimental results of dependence of (a) pre- and (b) post-FEC BER on OSNR in 100-km transmission obtained ML both Volterra processing in the conditions with both Tx and Rx nonlinearity induced (both NL), where three different cases for adaptive control of the coefficients of the Volterra filters at the first and seventh layers were evaluated.

the adaptive multi-layer filters including Volterra filter layers numerically and experimentally in the transmission of a 23 Gbaud PDM-64QAM signal over one span of a 100-km SMF. The results demonstrated that the adaptive multi-layer filters including Volterra filter layers could compensate for the nonlinearity that occurs in both Tx and Rx simultaneously and effectively when other impairments coexisted.

Chapter 6

Tx and Rx impairment compensation for ultra-long-haul single-mode fiber transmission by adaptive multi-layer filters with augmented inputs

The adaptive multi-layer filter architecture we investigated in the previous chapters enables efficient compensation of Tx and Rx impairments when other impairments such as CD exist. The filters in each layer can be designed to have minimum cross-terms and a filter length to compensate for the corresponding impairments, resulting in an efficient calculation of forward propagation. In terms of back propagation, the adaptive multi-layer filter architecture requires high computational resources, though it can be much mitigated for an embedded real-time DSP circuit by using offloaded external resources since back propagation or adaptive control for all the layers is not necessarily required to operate in real-time, as we discussed in Chapter 3.

However, the computational complexity of back propagation for the adaptive multi-layer architecture can be problematic even for external computational resources if applied to an ultra-long-haul SMF transmission. The adaptive multi-layer filters we investigated includes the layer for CD compensation, whose filter length can be very large in an ultra-long-haul SMF transmission since large CD is accumulated to the optical signal. In the back propagation of multi-layer filters, the temporal spread of the gradients progressively increases when going back through the layers because of the temporal spread of the filter response in each layer. It entails a large computational complexity for gradient calculation with back propagation through the layer for CD compensation with a large filter length.

In this chapter, we revisit the mutual non-commutativity of SL and WL filters and propose an adaptive multi-layer filter architecture to solve the abovementioned problem. In this architecture, CD compensation is performed on the received signal and its complex conjugate, and the two augmented signals are the inputs of 2×1 multi-input single-output (MISO) SL filters in the first layer, for Rx linear impairment compensation, of the adaptive multi-layer filters. In this multi-layer architecture, the CD compensation filters, whose coefficients can be static, are swept before the multi-layer filters, while the mutual non-commutativity of Rx impairment compensation and CD compensation is solved.

We evaluate the proposed adaptive multi-layer filters with augmented inputs through both numerical simulation and WDM transmission experiments of 16 channels of 32-Gbaud PDM-PCS-64QAM signals with an information rate (IR) of 2.4 b/sbl/pol in a 50-GHz grid over 10,200 km of SMF in which Tx and Rx IQ skews were digitally emulated. The results show that Tx IQ skew and Rx skew can be compensated by this adaptive multi-layer filter architecture with augmented inputs under a large CD accumulation over 10,000 km of SMF.

6.1 Background

As we discussed in the previous chapters, several approaches have been investigated to compensate for Rx linear impairments in optical fiber communication with CD accumulation. Independent CD compensation on I and Q signals can avoid mixing of IQ components in terms of Rx and a subsequent adaptive 4×2 MIMO filter enables Rx impairment compensation [9, 57, 58]. Alternatively, a large adaptive WL MIMO filter can compensate for Rx impairments, CD, and PMD/polarization state variation simultaneously [9]. As for Tx IQ impairments, they can be compensated by adaptive WL filters after most other impairments including CD, polarization effects, and carrier phase/frequency offset are compensated [54, 57, 58]. These previous works rely on individual adaptive filter blocks in which residual impairments degrade the performance.

In order to simultaneously compensate for Rx and Tx IQ impairments as well as polarization effects after independent CD compensation on I and Q signals, an adaptive 8×2 MIMO filter with phase offset compensation has been proposed [59, 60]. It uses an aggregated large adaptive filter that cannot efficiently compensate for multiple impairments having different characteristics.

In Chapter 3, we investigated an adaptive multi-layer SL&WL filter architecture whose filter coefficients in all layers can be controlled by gradient calculation with back propagation from the last layer and SGD. It enables precise compensation of Tx and Rx impairments and reasonable computational complexity. However, the computational complexity of back propagation can be problematic if applied to an ultra-long-haul SMF transmission since we have to calculate the gradients through the CD compensation filters with a large temporal spread in back propagation. To date, simultaneous compensation of Tx and Rx impairments in optical fiber communication systems has been demonstrated over a relatively short reach from 100 to 1500 km of SMF [57–60].

6.2 Adaptive multi-layer filters with augmented inputs and augmented CD compensation

To extend the adaptive multi-layer filter approach for simultaneous compensation of Tx and Rx impairments to ultra-long-haul optical fiber transmission, over a range such as 10,000 km of SMF in which the accumulated CD becomes large, we revisited the mutual non-commutativity of SL and WL filters in the adaptive multi-layer filters shown in Fig. 6.1. In this multi-layer filter architecture, any two impairments that are not mutually commutative are compensated in the reverse order of their occurrence. A WL filter and an SL filter are generally not mutually commutative.

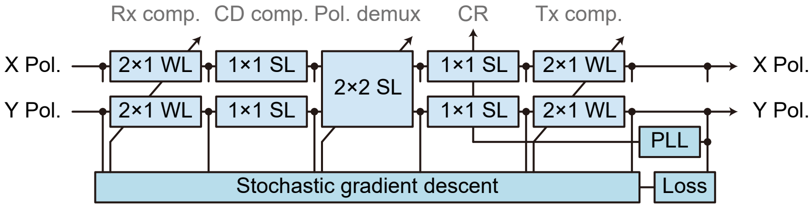


Figure 6.1. Adaptive multi-layer SL&WL filters to compensate for Tx and Rx impairments.

6.2.1 Equivalent architecture for cascaded WL and SL filters

We first examine the first and second layers of the adaptive SL&WL filters. The first layer consists of two 2×1 WL filters to compensate for Rx impairments. A 2×1 WL filter performs convolution of two filter responses to the input signal and its complex conjugate, and the output is a sum of the two signals after convolution. The term 2×1 indicates the structure of the filter coefficients. The second layer consists of two 1×1 SL filters with two polarizations to compensate for CD. Because both layers are independent in terms of two polarizations, we theoretically analyze one polarization here, without loss of generality. When a 2×1 WL filter and then a 1×1 SL filter for CD compensation operate on an input signal $x(t) \in \mathbb{C}$, the output of the 2×1 WL filter, $y(t) \in \mathbb{C}$, is given by

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau)x(t-\tau)d\tau + \int_{-\infty}^{\infty} h_{*1}(\tau)x^*(t-\tau)d\tau, \tag{6.1}$$

and the output of the 1×1 SL filter, $z(t) \in \mathbb{C}$, is given by

$$z(t) = \int_{-\infty}^{\infty} h_{\mathrm{CD}}(\tau)y(t-\tau)d\tau. \tag{6.2}$$

Here, h_1, h_{*1} represent the complex-valued impulse response of the 2×1 WL filter, and h_{CD} is the complex-valued impulse response of the 1×1 SL filter for CD compensation, which has a large temporal spread in ultra-long-haul SMF transmission. In addition, the superscript $*$ indicates the complex conjugate. By the distributive property, Eq. (6.2) can be rewritten as

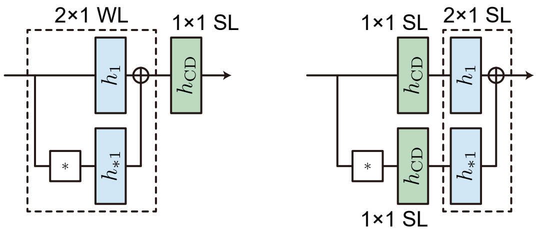
$$z(t) = \int_{-\infty}^{\infty} h_{\mathrm{CD}}(\tau)h_1(\tau')x(t{-}\tau{-}\tau')d\tau d\tau' + \int_{-\infty}^{\infty} h_{\mathrm{CD}}(\tau)h_{*1}(\tau')x^*(t{-}\tau{-}\tau')d\tau d\tau', \tag{6.3}$$

and then as

$$z(t) = \int_{-\infty}^{\infty} h_1(\tau') \left(\int_{-\infty}^{\infty} h_{\mathrm{CD}}(\tau)x(t-\tau'-\tau)d\tau \right) d\tau' + \int_{-\infty}^{\infty} h_{*1}(\tau') \left(\int h_{\mathrm{CD}}(\tau)x^*(t-\tau'-\tau)d\tau \right) d\tau'. \tag{6.4}$$

Consequently, the operation of a 2×1 WL filter and a 1×1 SL CD compensation filter on the input of x in this order is equivalent to the operation of a 2×1 SL MISO filter with a response of h_1, h_{*1} after CD compensation with a response of h_{CD} on x and x^*.

This behavior can also be understood from the block diagram shown in Fig. 6.2. Because of the mutual non-commutativity, simply changing the order of the 2×1 WL filter and the 1×1 SL filter, shown in Fig. 6.2(a), results in a different operation and does not work for simultaneous Rx IQ impairment and CD compensation. Instead, if the 1×1 SL filter for CD compensation is moved before summation via the distributive property, then the filters with complex-valued impulse responses h_1, h_{*1} and the one with h_{CD} are both 1×1 SL processes, accordingly they are mutually commutative and provide the equivalent configuration shown in Fig. 6.2(b)[1]. In optical fiber transmission systems, CD can be treated as static except in the case of optical path switching [3], and it is unnecessary to adaptively control the coefficients of CD compensation filters. Thus, by using the configuration shown in Fig. 6.2(b), Rx IQ impairment compensation can be achieved under CD, while CD compensation is swept before the adaptive filters and operated independently.



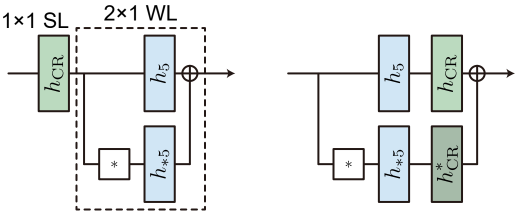
(a) (b)

Figure 6.2. DSP architecture for Rx impairment and CD compensation (a) by a 2×1 WL filter and a 1×1 SL filter, and (b) by an equivalent configuration.

---

[1] This exchange is possible since these processes are linear and 1×1. The exchange cannot be applied to nonlinear processes.

We can apply a similar approach to the fourth and fifth layers of the adaptive SL&WL filters. The operation of a 1×1 SL filter to compensate for carrier phase and frequency offset and then a 2×1 WL filter to compensate for Tx IQ impairments, as shown in Fig. 6.3(a), is equivalent to the configuration shown in Fig. 6.3(b). This explains the embedded carrier phase offset compensation in an 8×2 WL MIMO filter [59].



(a) (b)

Figure 6.3. DSP architecture for carrier phase offset and Tx impairment compensation (a) by a 1×1 SL filter and a 2×1 WL filter, and (b) by an equivalent configuration.

6.2.2 Adaptive multi-layer filter architecture for efficient back propagation

Figure 6.4 shows the adaptive multi-layer filter architecture, based on the above analysis, with augmented inputs consisting of the received signal and its complex conjugate for CD compensation. In this architecture, the CD compensation operates on the signal and its complex conjugate before the multi-layer filters. Then, with the augmented inputs consisting of the signal and its complex conjugate after CD compensation, the adaptive multi-layer filters compensate for Rx impairment, PMD/polarization state variation, carrier phase/frequency offset, and Tx impairment. The first layer of the multi-layer filters consists of two 2×1 SL MISO filters with two polarizations to compensate for Rx impairment. The subsequent layers are the same as in our previous adaptive multi-layer SL&WL filters: the second layer consists of a 2×2 SL MIMO filter for polarization demultiplexing; the third layer consists of two 1-tap 1×1 SL filters with two polarizations for carrier recovery; and the fourth layer consists of two 2×1 WL filters with two polarizations to compensate for Tx impairment. In this investigation, these filters are half-symbol-spaced FIR filters. The filter coefficients in the first, second, and fourth layers are all adaptively controlled by gradient calculation with back propagation and SGD to minimize the loss function, which is composed of the last layer's outputs. The coefficients in the third layer for carrier recovery are controlled by a PLL.

As mentioned above, in the architecture shown in Fig. 6.4, the CD compensation operates independently before the adaptive multi-layer filters. CD compensation filters require about 5600 half-symbol-spaced taps for transmis-

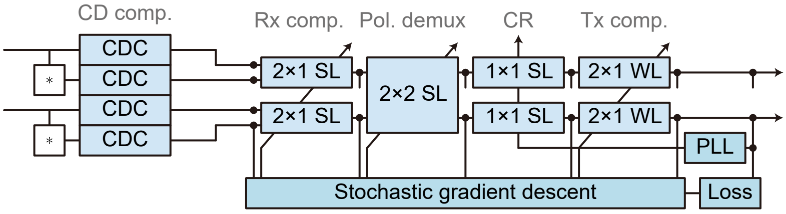


Figure 6.4. Architecture of the adaptive multi-layer filters with augmented inputs of the signal and its complex conjugate with CD compensation to compensate for Tx and Rx impairments under a large CD accumulation.

sion over 10,000 km of conventional SMF having a dispersion coefficient of 17 ps/nm/km with a 32 Gbaud signal [5]. Here, we mitigate the computational complexity of gradient calculation by avoiding back propagation through the CD compensation filters, which have a large temporal spread. In the experiments described later, the tap lengths of the filters for Rx/Tx IQ impairment compensation and for polarization demultiplexing were 5 and 21, respectively. With these tap lengths, the required number of complex multiplications for back propagation of the adaptive multi-layer filters with augmented inputs is reduced to 0.2% of that required by our previous adaptive ML SL&WL filters including CD compensation. As for forward propagation, although this architecture doubles the number of CD compensation filters[2], they can be implemented in the frequency domain to the reduce the computational complexity [61], which is very effective, especially for filters with a large temporal spread.

It is worth mentioning that applying CD compensation to both the signal and its complex conjugate is not equivalent to applying CD compensation to the signal and then obtaining its complex conjugate. Rather, inputting the augmented signal and its complex conjugate with CD compensation to the adaptive filters is equivalent to inputting independently CD-compensated IQ signals [9]. The augmented signals $y(t), y_c(t) \in \mathbb{C}$ after the operation of CD compensation on the received signal $x(t)$ and its complex conjugate are represented as

$$\begin{pmatrix} y(t) \\ y_c(t) \end{pmatrix} = \int_{-\infty}^{\infty} h_{\mathrm{CD}}(\tau) \begin{pmatrix} x(t-\tau) \\ x^*(t-\tau) \end{pmatrix} d\tau. \tag{6.5}$$

Meanwhile, consider the case of CD compensation operating independently on the IQ components $x_{\mathrm{I}}(t), x_{\mathrm{Q}}(t) \in \mathbb{R}$ of the received signal $x(t) = x_{\mathrm{I}}(t) + ix_{\mathrm{Q}}(t)$. In this case, the obtained signals $y_i(t), y_q(t) \in \mathbb{C}$ are represented as

$$\begin{pmatrix} y_i(t) \\ y_q(t) \end{pmatrix} = \int_{-\infty}^{\infty} h_{\mathrm{CD}}(\tau) \begin{pmatrix} x_{\mathrm{I}}(t-\tau) \\ x_{\mathrm{Q}}(t-\tau) \end{pmatrix} d\tau. \tag{6.6}$$

---

[2] The multi-layer filter architecture investigated in Chapter 2 is more efficient in terms of the computational complexity for forward propagation.

Here, the following relation holds [42]:

$$\begin{pmatrix} x(t) \\ x^*(t) \end{pmatrix} = T \begin{pmatrix} x_{\mathrm{I}}(t) \\ x_{\mathrm{Q}}(t) \end{pmatrix}, \tag{6.7}$$

with

$$T = \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}. \tag{6.8}$$

Here, T satisfies $TT^\dagger = T^\dagger T = 2I$, where I is the identity matrix and † denotes the Hermitian conjugate. As a result, the following relation also holds:

$$\begin{pmatrix} y(t) \\ y_c(t) \end{pmatrix} = T \begin{pmatrix} y_i(t) \\ y_q(t) \end{pmatrix}. \tag{6.9}$$

In consequence, the input of $y(t)$ and $y_c(t)$ to the adaptive filters can yield the same results as the input of $y_i(t)$ and $y_q(t)$, with only an additional transformation T.

As shown in Fig. 6.2, the coefficients of the 2×1 SL MISO filter in the first layer of the adaptive multi-layer filters straightforwardly correspond to those of the 2×1 WL filter for Rx impairment compensation. Moreover, the 2×1 WL filter for Tx IQ impairment compensation in the fourth layer is the same as in the previous multi-layer SL&WL filters. Thus, the IQ impairment monitoring via the obtained adaptive filter coefficients in the corresponding layers investigated Chapter 4 is immediately applicable.

6.3 Evaluation in simulation

We first evaluated the adaptive multi-layer filter architecture described above in simulations using a simple model. We adopted PCS[3], which is regarded as a promising technique because it enables high sensitivity and precise rate adaptability for the received SNR from short-reach to long-haul transmission [11, 112, 113] in this simulation. We used a 32-Gbaud PDM-PCS-64QAM signal with an IR of 1.6 b/sbl/pol, which corresponds to the IR of QPSK when FEC is applied with the same code rate. CD corresponding to that of 10,000 km of SMF was added to the signal. We focused on IQ skew as the source of Tx and Rx impairments.

Simulation model

On the Tx side, a 32-Gbaud PDM-PCS-64QAM signal with an IR of 1.6 b/sbl/pol was generated. Three frames of low-density parity check FEC for DVB-S2 with a frame length of 64,800 and a code rate of 0.8, obtained by loading random bits in the payload, were generated as the transmitted data for the IQ components and two polarizations, which resulted in a total of 12 frames. A distribution matcher with the probabilistic amplitude shaping architecture [112] and constant composition distribution matching (CCDM) [114] mapped the data to

---

[3] PCS often uses high-order QAM as a base format and optimizes the probability of occurrence of symbol points for a transmission channel, usually according to a Maxwell-Boltzmann distribution for an additive white Gaussian noise channel. Remarkable long-haul transmission capability has been achieved with PCS [109–111].

the PDM-PCS-64QAM signal. The probability distribution was an Maxwell-Boltzmann distribution, and the IR was set to 1.6 b/sbl/pol. To enable use of the pilot-based [70, 71] data-aided LMS algorithm for coefficient control of the adaptive filters, one pilot symbol was inserted every 15 symbols. The pilot symbols were randomly sampled from the generated PDM-PCS-64QAM signal in the same manner so as to match the statistical characteristics of the pilot symbols and the transmitted signal. To enable the Rx DSP to detect the timing of the pilot symbol sequence's head, a preamble of 64 symbols was inserted in every FEC frame. The preamble symbols consisted of a random sequence from the inner four symbol points of the 64QAM, which enabled the Rx DSP to detect the preamble on the basis of the signal power. In addition, to match the limitation of an instrument that was used in the experiment described later, dummy data consisting of 448 symbols that were randomly sampled from the PDM-PCS-64QAM signal was added after each preamble.

The resulting signal with the pilot and preamble symbols was oversampled to 32-fold oversampling and shaped with a root-raised cosine filter having a roll-off factor of 0.1. The four IQ components with two polarizations corresponded to the electrical signals that a DAC would output. Tx IQ skew was emulated for these signals. Lastly, a laser source with a frequency of 193.3 THz and a linewidth of 100 kHz was modulated by these signals, and a 32 Gbaud PDM-PCS-64QAM signal was thus generated.

Next, CD of 170 ns/nm, which corresponds to 10,000 km of SMF, was added to the generated signal. After adding random polarization rotation, the signal's OSNR was set to 15 dB/0.1 nm. The optical signal was received through coherent detection with an LO source having a linewidth of 100 kHz, no frequency offset, and low-pass filters whose 3-dB bandwidth was 0.8 times the symbol rate. Rx IQ skew was then emulated for the four IQ components with two polarizations, which corresponded to the electrical signals that would be input to an ADC. The ADC sampled these signals with two-fold oversampling.

Next, DSP was performed. Specifically, low-pass filtering with a 3-dB bandwidth equal to the symbol rate was performed on the four received signals for resampling, and power normalization was performed individually on the IQ components with two polarizations. The signals were resampled to two-fold oversampling on the basis of the timing error [86], and matched filtering with the root-raised cosine filter was applied. Then, the adaptive multi-layer filters shown in Fig. 6.4 were applied. The CD compensation on the signal and its complex conjugate operated in the frequency domain.

The adaptive multi-layer filters, with augmented inputs of the received signal and its complex conjugate with CD compensation, consisted of four layers as described in the previous section. This DSP configuration is referred to as ML. The tap length of the 2×1 SL MISO filters in the first layer, for Rx impairment compensation, was set to five. The tap lengths of the 2×2 SL MIMO filter in the second layer, for polarization demultiplexing, and the 2×1 WL filters in the fourth layer, for Tx impairment compensation, were set to 21 and 5, respectively. The coefficients for these filters were initialized as one at the center of the main diagonal, with the remaining values set to zero. The filter coefficients in the first, second, and fourth layers were adaptively updated with back propagation and SGD by the data-aided LMS algorithm at the timings of the preamble, dummy, and pilot symbols, with no update at the data symbol timings. The step sizes were 10^{-2} for the first and fourth layers and 5×10^{-2} for the second layer. The coefficients of the third layer, for carrier phase and frequency offset compensation, were controlled by a PLL from the last layer's outputs. Here, the phase error was detected from data symbols at the timings of the preamble, dummy, and pilot symbols and from decision-directed symbols at other timings.

For timing alignment with the known pilot sequence, the following procedure was carried out before applying the adaptive multi-layer filters. The preamble timing was roughly detected from the received signal after CD compensation on the basis of a moving average of the sample power. Then, for about ± 10 symbols around the detected timing, adaptive control of the multi-layer filters were attempted over a duration of about 10,000 symbols. For this adaptive control, the tap length of the 2×2 SL MIMO filter was set to one, and the precise timing of the pilot sequence's head was detected on the basis of the magnitude of the loss after adaptive control. After the precise timing alignment, the received signal and its complex conjugate with CD compensation were input to the multi-layer filters, and the pilot-based adaptive control was performed again. From the symbol-spaced outputs of the adaptive multi-layer filters after convergence of the filter coefficients, the preamble, dummy, and pilot symbols were removed, and FEC and CCDM decoding were then performed. A sum-product algorithm with maximum 50 iterations was used for FEC decoding. The post-decoding BER and the NGMI [74] averaged over two polarizations were evaluated.

To provide a reference for conventional DSP, we also evaluated a configuration in which a 2×2 SL MIMO filter embedding carrier recovery [115] for polarization demultiplexing and carrier phase and frequency offset compensation was applied after CD compensation. This is referred to hereafter as the 2×2 SL configuration[4]. The coefficients of the 2×2 SL MIMO filter were controlled by the pilot-based data-aided LMS algorithm, and carrier recovery was controlled by a PLL. The step size was 5×10^{-2}.

Dependence of IQ skew impact on IR of PCS-QAM

In an ultra-long-haul transmission, the received OSNR tends to be low, and thus, PCS-QAM with a higher IR is not applicable. Before evaluating the performance of Tx and Rx impairment compensation in an ultra-long-haul SMF transmission with the adaptive multi-layer filters with the augmented inputs, we evaluate the dependence of the impact of IQ skew on the IR of PCS-64QAM through simulation by using the reference 2×2 SL configuration. In this simulation, the accumulated CD was set to 17 ns/nm, which corresponds to 1000 km of SMF transmission, and different received OSNR values was used. Tx or Rx IQ skew was again emulated in the X polarization.

Figure 6.5 shows the NGMI results for reception of a 32-Gbaud PDM-64QAM signal with several different IRs by the 2×2 SL configuration while the Tx or Rx IQ skew was varied from -31 ps to +31 ps. The received OSNR was 30 dB/0.1 nm. The optimum step size for stochastic gradient descent depended on the PCS-QAM IR, and it was roughly optimized. In the case of PCS-64QAM with an IR of 4.8 b/sbl/pol, which corresponds to the IR of 64QAM when FEC is applied with the same code rate, an NGMI above 0.85 was achieved within ±4 ps of Tx IQ skew, whereas the range expanded to about ±22 ps with an IR of 1.6 b/sbl/pol. As for Rx IQ skew, an NGMI above 0.85 was achieved within

---

[4] This configuration is the same as used in the previous chapters as a reference.

±4 ps with an IR of 4.8 b/sbl/pol, and within about ±26 ps with an IR of 1.6 b/sbl/pol. These results show that the tolerance to IQ skew increases as the PCS-QAM IR decreases, even in the same base QAM format.

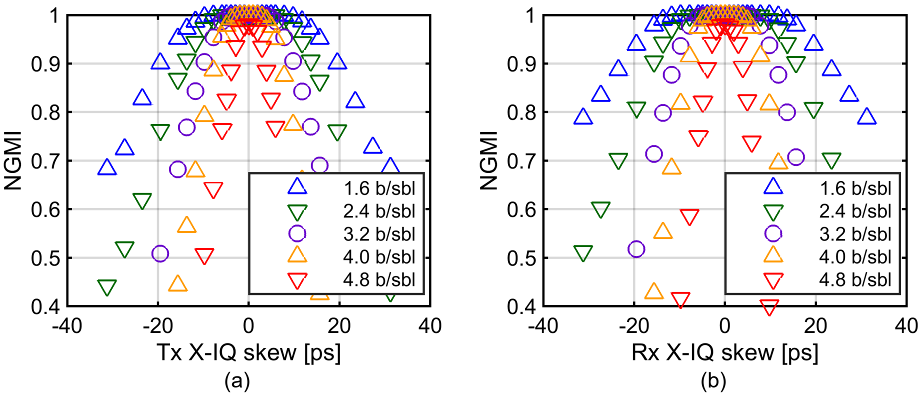


Figure 6.5. Simulated NGMI results for reception of a 32-Gbaud PDM-64QAM signal with several IRs by the 2×2 SL configuration when (a) the Tx X-IQ skew or (b) the Rx X-IQ skew was varied.

Next, Fig. 6.6 shows the NGMI results for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol and several differenct received OSNRs by the 2×2 SL configuration while the Tx or Rx IQ skew was varied. For several of the received OSNRs, the NGMI degradation due to IQ skew showed a similar behavior. However, regardless of the IQ skew, the NGMI deteriorated as the received OSNR decreased. This caused the region in which the NGMI was above a certain threshold to become narrow when the received OSNR was low after all, which indicates that precise Tx and Rx IQ impairment compensation is still important for ultra-long-haul transmission.

Tx and Rx IQ skew compensation

Figure 6.7 shows the resulting constellations after the ML configuration or the 2×2 SL configuration under several conditions: without both Tx and Rx IQ skew, with Tx IQ skew of 16 ps, with Rx IQ skew of 16 ps, and with both Tx and Rx IQ skews of 16 ps. In this evaluation, the Tx or Rx IQ skew was emulated as the delay of the Q component in the X polarization. The constellations for the X polarization are shown in the figure. When Tx or Rx IQ skew of 16 ps was emulated, a clear PCS-64QAM constellation like in the case without both Tx and Rx IQ skew could not be obtained by the 2×2 SL configuration. In contrast, with Tx or Rx IQ skew, the ML configuration obtained similar constellations to the case without both Tx and Rx IQ skew. Moreover, even in the case with both Tx and Rx skews of 16 ps, the adaptive multi-layer filters still obtained a similar constellation.

Figure 6.8 shows the NGMI after the 2×2 SL configuration or the ML configuration when the Tx or Rx IQ skew in the X polarization was swept from -31 to

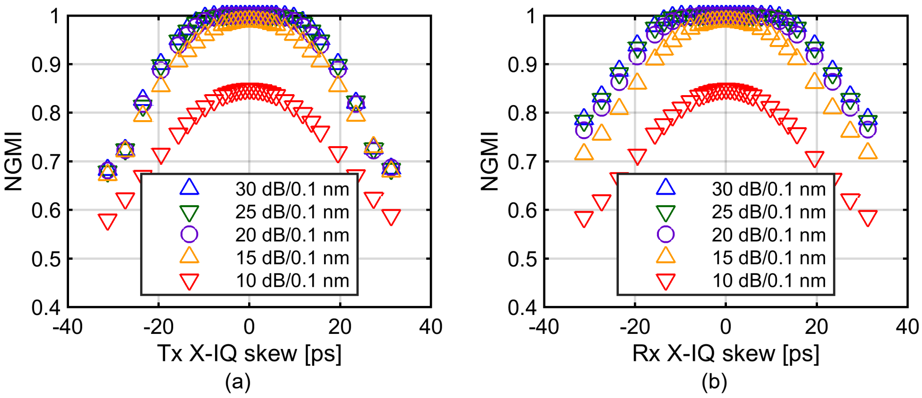


Figure 6.6. Simulated NGMI results for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol and several received OSNRs by the 2×2 SL configuration when (a) the Tx X-IQ skew or (b) the Rx X-IQ skew was varied.

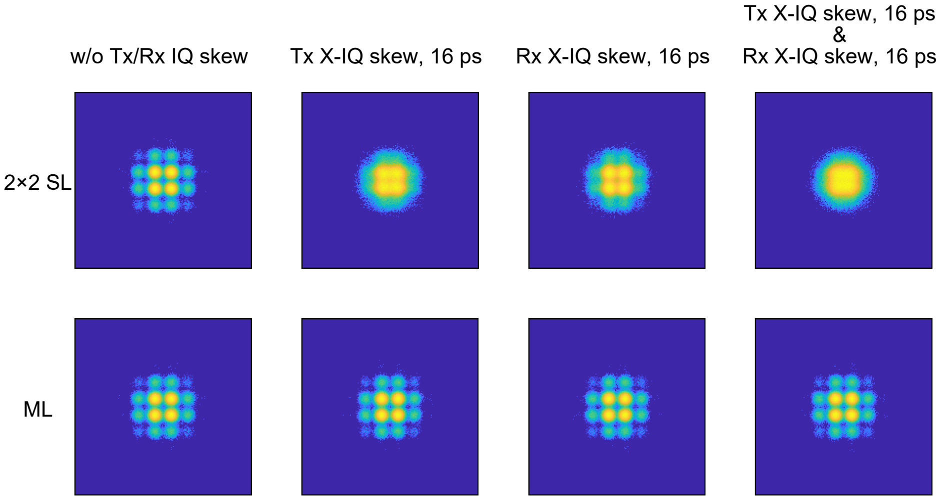


Figure 6.7. Constellations of the X polarization for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol by the 2×2 SL configuration or the ML configuration, under several Tx/Rx IQ skew conditions.

+31 ps, and Fig. 6.9 shows the post-decoding BER under the same conditions. The error-free BER results are plotted at a level of 10^{-5} for readability. In the case of the 2×2 SL configuration, which could not compensate for Tx or Tx IQ impairments, both the Tx and Rx IQ skews degraded the NGMI. Moreover, the post-decoding BER was not error-free when the Tx IQ skew was beyond ±14 ps, and when the Rx IQ skew was beyond ± 20 ps. In contrast, with the adaptive multi-layer filters, NGMI results similar to those in the case without IQ skew and error-free post-decoding BER results were achieved with up to ±23 ps of Tx IQ skew, and with over ±31 ps of Rx IQ skew. Both figures also show the results obtained by the adaptive multi-layer filters when the tap lengths of the filters for Rx and Tx IQ impairment compensation in the first and fourth layers were nine. These results show that the ML configuration' tolerance to Tx IQ skew did not increase when the tap length of the filters for Tx IQ impairment compensation increased to nine. After several evaluations, the PLL that determined the filter coefficients of the third layer became unstable when the Tx IQ impairment was large. This behavior will be a future issue for us to resolve, although Tx IQ skew compensation over ±10 ps will still be useful in practice for current optical fiber communication systems.

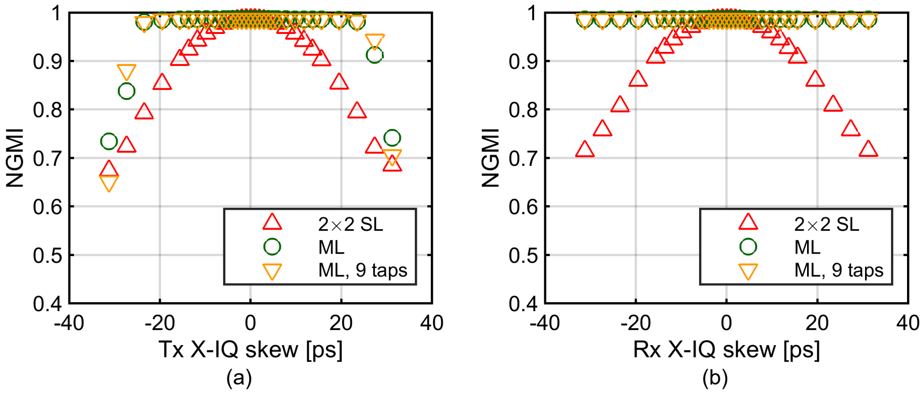
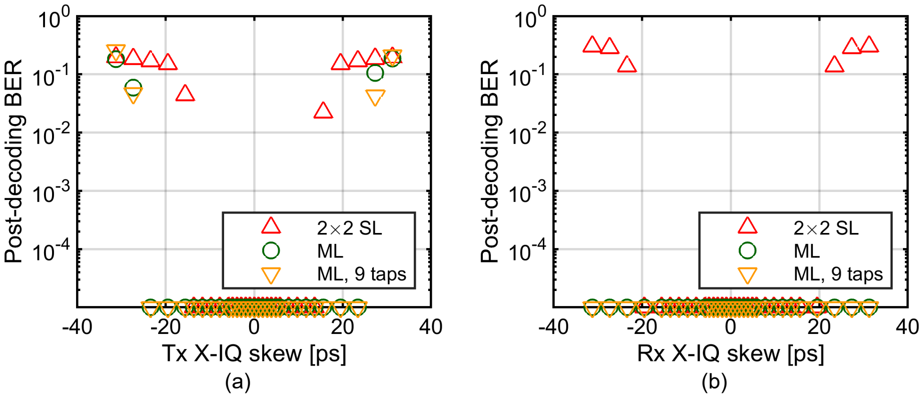


Figure 6.8. NGMI for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol by the 2×2 SL configuration or the ML configuration, with a tap length of five or nine taps for Tx and Rx impairment compensation. The results are shown for variation of (a) the Tx IQ skew in the X polarization and (b) the Rx IQ skew in the X polarization.

We also evaluated the capability for simultaneous compensation of Tx and Rx IQ skews. Random Tx and Rx skews in two polarizations from a uniform distribution within ±16 ps were emulated simultaneously, with a total of 1000 random realizations for evaluation. A histogram of the NGMI in the cases with the 2×2 SL configuration and the ML configuration is shown in Fig. 6.10. In the 2×2 SL case, the NGMI distribution was spread to lower values by the Tx and Rx IQ skews. In contrast, the ML configuration achieved a stable NGMI of around one. These results confirm that the adaptive multi-layer filters with augmented inputs could sufficiently and simultaneously compensate for Tx and



(a)

(b)

Figure 6.9. Post-decoding BER for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol by the 2×2 SL configuration or the ML configuration, with a tap length of five or nine taps for Tx and Rx IQ impairment compensation, when (a) the Tx X-IQ skew and (b) the Rx X-IQ skew was varied.

Rx IQ skews under a large CD accumulation corresponding to 10,000 km of SMF.

Convergence of loss

Figure 6.11 shows examples of the time evolution of the loss with the ML configuration under 10 random realizations of Tx and Rx IQ skew. The loss at the timings of the preamble, dummy, and pilot symbols was taken with a moving average over 10 symbols. The results show that the adaptive control converged within a few thousand symbols.

Tx and Rx impairment monitoring

We also monitored the Tx and Rx IQ skews from the filter coefficients of the fourth and first layers in the ML configuration after convergence. Fig. 6.12 shows the monitored values from the corresponding layers for 1000 random realizations in which the Tx and Rx IQ skews were simultaneously emulated in two polarizations. For all of the Tx and Rx IQ skews in two polarizations, the monitored values agreed well with the emulated skews, as the error was within ±1 ps. Thus, we confirmed that impairment monitoring via the adaptive filter coefficients works for the adaptive multi-layer filters with augmented inputs under a large CD accumulation.

Influence of residual CD

In the ML configuration shown in Fig. 6.4, CD compensation is applied before the adaptive multi-layer filters. In practical optical fiber transmission systems, the accumulated CD in the received signal is estimated on the Rx side [3]. When

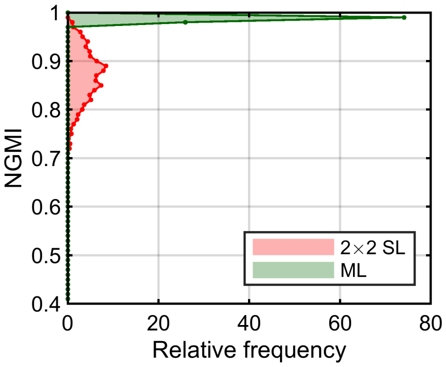


Figure 6.10. Histogram of the NGMI for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol by the 2×2 SL configuration or the ML configuration under 1000 random realizations of simultaneous Tx and Rx IQ skew in two polarizations.

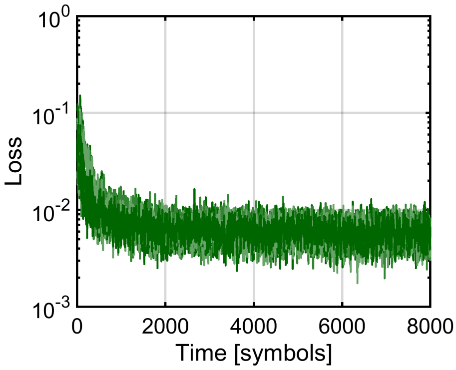
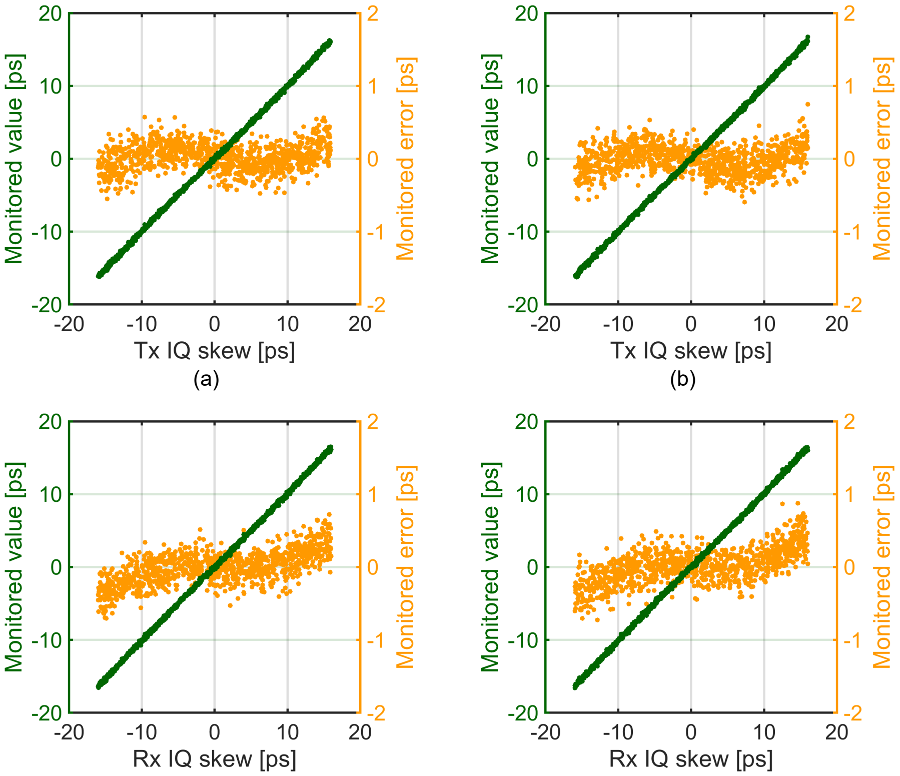


Figure 6.11. Convergence of the loss at the pilot symbol timings with the ML configuration. Examples of the time evolution are plotted for 10 random realizations of simultaneous Tx and Rx IQ skew.

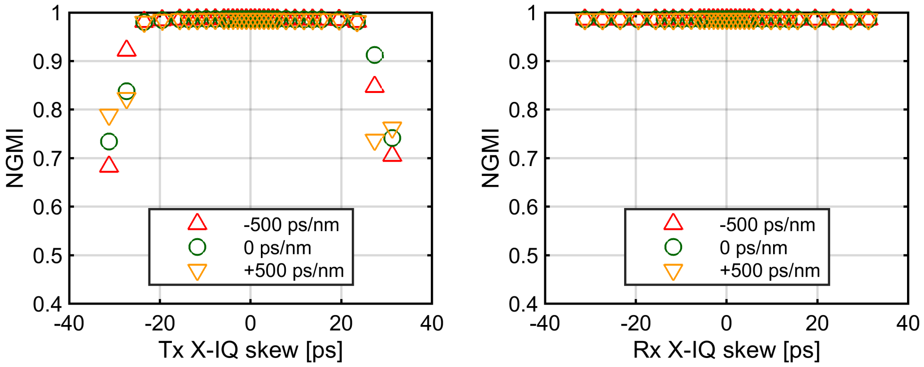


(a) (b) (c) (d)

Figure 6.12. Impairment monitoring via the coefficients of the ML configuration for (a) Tx IQ skew in the X polarization, (b) Tx IQ skew in the Y polarization, (c) Rx IQ skew in the X polarization, and (d) Rx IQ skew in the Y polarization, under 1000 random realizations of simultaneous Tx and Rx IQ skew.

such CD estimation is not fully accurate, the residual CD after CD compensation can also be compensated in a subsequent adaptive 2×2 SL MIMO filter for polarization demultiplexing in a conventional DSP configuration. Here, we evaluated the influence of such residual CD on the ML configuration with augmented inputs consisting of the received signal and its complex conjugate with CD compensation. In this simulation, the compensated CD value for the signal and its complex conjugate before the adaptive multi-layer filters did not precisely match the true value.

Figure 6.13 shows the NGMI when the Tx or Rx IQ skew in the X polarization was swept. The compensated CD value for the signal and its complex conjugate before the adaptive multi-layer filters was set to ±500 ps/nm and zero, which can be compensated by the 2×2 SL MIMO filter in the second layer, from the true value of 170 ns/nm. When the residual CD was ±500 ps/nm, an NGMI similar to that in the case without residual CD was obtained, regardless of the existence of Tx and Rx IQ skew. Thus, the Tx and Rx IQ impairment compensation by the adaptive multi-layer filters with augmented inputs can work if the residual CD after CD compensation and before the multi-layer filters is within the effective compensation range of the second layer in the multi-layer filters.



(a)

(b)

Figure 6.13. Simulated NGMI results for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol by the ML configuration when the compensated CD value before the adaptive multi-layer filters did not precisely match the true value, while varying (a) the Tx X-IQ skew or (b) the Rx X-IQ skew.

Compensation of IQ imbalance and phase deviation in Tx and Rx

We focused on IQ skew as the source of Tx and Rx IQ impairments in the previous evaluations. A WL filter can compensate for other linear impairments that occur in Tx and Rx, such as IQ imbalance and phase deviation, as we discussed in Chapter 4. Here, we evaluated compensation of IQ imbalance and phase deviation by the adaptive multi-layer filters with augmented inputs consisting of the received signal and its complex conjugate with CD compensation. The simulation model was the same as before, except that IQ imbalance and

IQ phase deviation were emulated in Tx and Rx. In DSP, power normalization was performed on the complex-valued received signal, rather than performed individually on the IQ components, to evaluate compensation of IQ imbalance.

Figure 6.14 shows the results of constellations by the 2×2 SL configuration or the ML configuration under Tx/Rx IQ imbalance and phase deviation. IQ imbalance of 30% and phase deviation of 10° were emulated in X polarization in Tx/Rx. The constellations for the X polarization are shown in the figure. In the case of the 2×2 SL configuration, all Tx/Rx IQ imbalance and phase deviation distorted the constellation, though a little effect of Rx phase deviation of 10° in X polarization was observed. In contrast, the ML configuration obtained similar constellations to the case without IQ impairments shown in Fig. 6.14. Since Tx IQ imbalance distorted added Gaussian noise, the observed constellation looked slightly different. However, the ML configuration could compensate for Tx IQ imbalance, compared to the case of the 2×2 SL configuration. Therefore, we confirmed that the adaptive multi-layer filters with augmented inputs can also compensate for IQ imbalance and phase deviation in Tx/Rx.

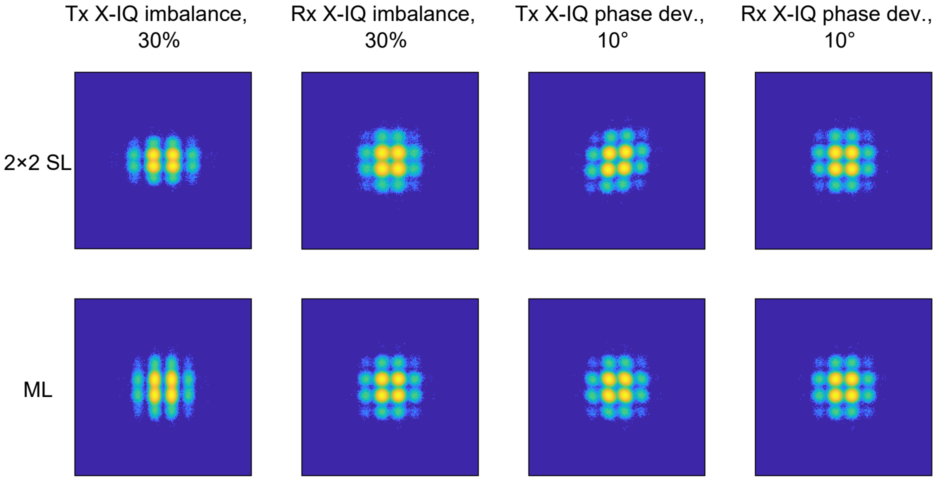


Figure 6.14. Simulated results of constellations of the X polarization for reception of a 32-Gbaud PDM-64QAM signal with an IR of 1.6 b/sbl/pol by the 2×2 SL configuration or the ML configuration under Tx/Rx IQ imbalance and phase deviation.

6.4 Evaluation in transmission experiment

We evaluated the adaptive multi-layer filter architecture with augmented inputs of the received signal and its complex conjugate with CD compensation in a transmission experiment over 10,000 km of SMF. In this experiment, we transmitted 16-channel WDM signals of 32-Gbaud PDM-PCS-64QAM with an

IR of 2.4 b/sbl/pol, which corresponds to the IR of 8QAM when FEC is applied with the same code rate. The signals were transmitted over a recirculating loop configuration consisting of five spans of 60-km pure-silica-core (PSC) SMF. As in the simulation described above, we focused on IQ skew as the source of Tx and Rx impairments.

Experimental setup

Figure 6.15 shows a schematic diagram of the experimental setup. On the Tx side, 32-Gbaud PDM-PCS-64QAM signals were generated, with frequencies ranging from 192.90 to 193.65 THz in a 50-GHz grid. The signal at a frequency of 193.3 THz, which was received and evaluated, was generated by modulating a laser source having a linewidth of about 100 kHz with the outputs of a four-channel DAC at a sampling rate of 92 GS/s and a vertical resolution of eight bits. The transmitted data were generated by the same procedure that was described above for the simulation. The IR of the PDM-PCS-64QAM signal was set to 2.4 b/sbl/pol. The signal, with pilot and preamble symbols included, was oversampled to two-fold oversampling and shaped with a root-raised cosine filter having a roll-off factor of 0.1. Tx IQ skew was digitally emulated in the X polarization as a delay of the Q component. A DAC outputted the four IQ components with two polarizations after resampling to 92 GS/s. The 15 remaining channels were generated by using 15 channels of laser sources combined with a polarization-maintaining AWG and a four-channel DAC at a sampling rate of 64 GS/s without IQ skew. The signal under evaluation and the other 15 channels were then combined after power equalization by a WSS, and low-speed polarization scrambling ($10\times2\pi$ rad/s) was performed.

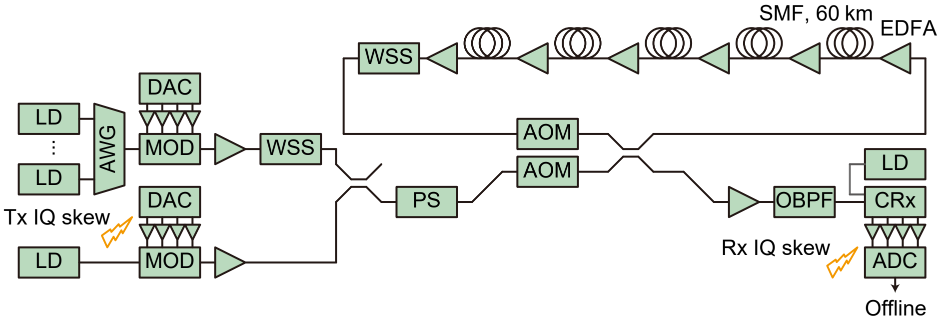


Figure 6.15. Experimental setup for WDM transmission of 32-Gbaud PDM-PCS-64QAM signals with an IR of 2.4 b/sbl/pol over a loop configuration consisting of five spans of 60-km SMF. AWG: arrayed waveguide grating, WSS: wavelength selective switch, AOM: acousto-optic modulator.

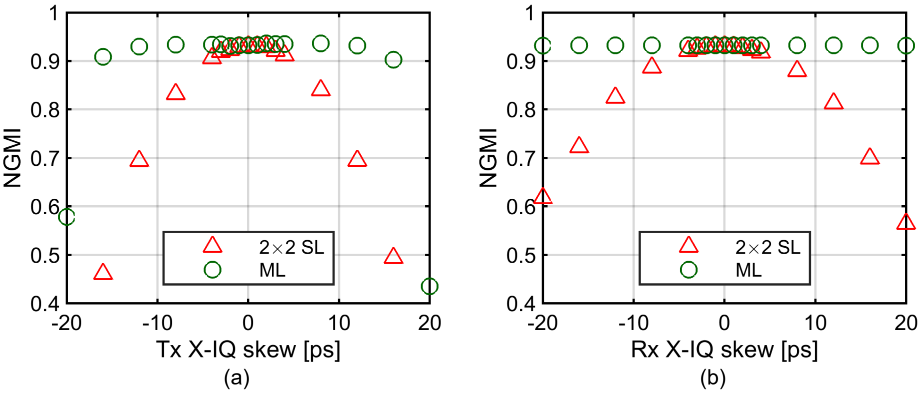
The generated WDM signals were transmitted to the transmission line through an acousto-optic loop switch. As mentioned above, the transmission line had a loop configuration composed of five spans of 60-km PSC SMF, with EDFAs. The dispersion coefficient and effective core area of the SMF were about 21 ps/nm/km and 153 μm^2, respectively. The span input power was -1 dBm/ch, which was an optimized value in a preliminary experiment. After five spans, the nonuniform loss and gain in the loop for the WDM channels were equalized by the WSS. After 34 loops, or 10,200 km, the OSNR was 18.7 dB/0.1 nm.

After transmission over the loop configuration, the evaluated signal was demultiplexed by an optical band-pass filter with a 3-dB bandwidth of 50 GHz, and then received by a polarization-diversity coherent receiver. A laser source with a linewidth of about 100 kHz was used as an LO, which was free-running with the Tx laser source and had an average frequency offset of about 70 MHz. The four outputs of the coherent receiver were sampled by a four-channel oscilloscope, which was used as an ADC, at a sampling rate of 80 GS/s and a vertical resolution of eight bits. A received signal consisting of 2 MS of successive data was obtained three times under each condition. Rx IQ skew was digitally emulated in the X polarization after sampling as a delay of the Q component. The initial Tx and Rx IQ skews in the experimental setup when IQ skew was not emulated were calibrated as closely to zero as possible. After emulating the Rx IQ skew, DSP was performed offline. We evaluated the same configurations that were described above for the simulation, namely, the adaptive ML filters with the augmented inputs (ML) and the reference (2×2 SL).

Tx and Rx IQ skew compensation

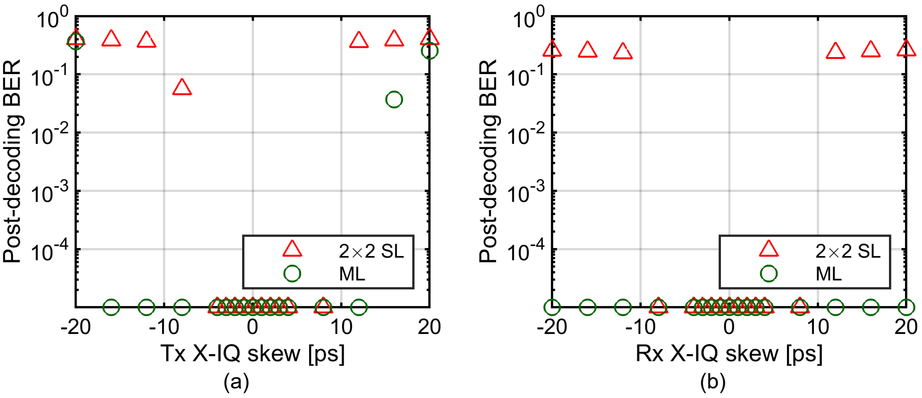
Under the condition of transmission over 34 loops, or 10,200 km, the Tx or Rx IQ skew was varied from -20 to +20 ps. Figure 6.16 shows the NGMI after the 2×2 SL configuration or the adaptive ML filters, and Fig. 6.17 shows the post-decoding BER. The results were averaged over the three received waveforms that were obtained. With the 2×2 SL configuration, the post-decoding BER was error-free only within ±4 ps of Tx IQ skew and within ±8 ps of Rx IQ skew, whereas post-decoding errors occurred outside those ranges. In contrast, with the ML configuration, error-free post-decoding BER results and NGMI results similar to those in the case without IQ skew were achieved with up to ±12 ps of Tx IQ skew, and with over ±20 ps of Rx IQ skew. As was observed in the simulation, when the Tx IQ skew was around ±20 ps, the NGMI and post-decoding BER deteriorated with the adaptive ML filters. Except for that case, however, these results experimentally confirm that the ML configuration could compensate for both Tx and Rx IQ skews under a large CD in ultra-long-haul transmission.

Next, we evaluated simultaneous compensation of the Tx and Rx IQ skews for transmission over 10,200 km. Figure 6.18 shows the NGMI after the 2×2 SL configuration and the ML configuration when Tx and Rx IQ skews in the X polarization were simultaneously emulated. With the 2×2 SL configuration, as shown in Fig. 6.18(a), both the Tx and Rx IQ skews degraded the NGMI, although the degradation was not smooth when the Tx and Rx skews became large. The reason was that the adaptive 2×2 SL MIMO filter could not converge stably in this region. The NGMI variation was relatively smaller when the Rx IQ skew was varied than when the Tx IQ skew was varied in this region, because the Rx IQ skew was emulated for the same three received waveforms that were obtained for a certain emulated Tx IQ skew. In contrast, except in the region around a Tx IQ skew of ±20 ps, a similar NGMI was achieved with the ML configurations even when the Tx and Rx IQ skews were simultaneously



(a) (b)

Figure 6.16. NGMI for reception of a 32-Gbaud PDM-64QAM signal with an IR of 2.4 b/sbl/pol by the 2×2 SL configuration or the ML configuration after 10,200 km of SMF transmission. The results were obtained while emulating (a) Tx IQ skew in the X polarization or (b) Rx IQ skew in the X polarization.



(a) (b)

Figure 6.17. Post-decoding BER for reception of a 32-Gbaud PDM-64QAM signal with an IR of 2.4 b/sbl/pol by the 2×2 SL configuration or the ML configuration after 10,200 km of SMF transmission, with emulation of (a) Tx IQ skew in the X polarization or (b) Rx IQ skew in the X polarization.

emulated.

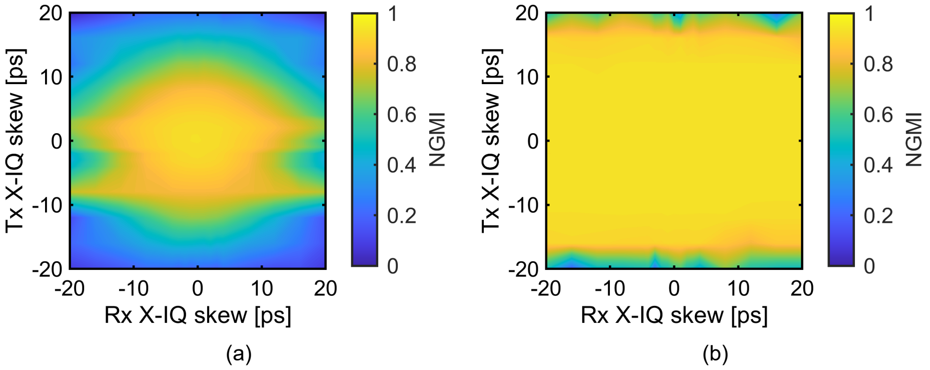


Figure 6.18. NGMI for reception of a 32-Gbaud PDM-64QAM signal with an IR of 2.4 b/sbl/pol by (a) the 2×2 SL configuration or (b) the ML configuration, after 10,200 km of SMF transmission when Tx and Rx IQ skew in the X polarization were simultaneously emulated.

Tx and Rx impairment monitoring

We also evaluated impairment monitoring via the adaptive multi-layer filter coefficients in the corresponding layers. For 13 conditions each of Tx and Rx IQ skews of 0, ±1, . . . , ±4, ±8, ±12 ps, giving a total of 169 conditions, we obtained three results under each condition. Figure 6.19 shows the monitoring results for Tx and Rx IQ skews in the X polarization, which show that the monitored values agreed well with the emulated skews. The monitored values for the Rx IQ skew showed a relatively large error, though the maximum error was 1.1 ps. Consequently, these results experimentally confirm that IQ impairment in both Tx and Rx could be monitored in ultra-long-haul transmission via the coefficients of the adaptive multi-layer filters with augmented inputs.

Transmission performance of PCS with adaptive multi-layer filters with augmented inputs

Finally, we evaluated the optimization of the PCS-QAM signal's IR with fixed FEC, for transmission over 10,200 km, by the adaptive ML filters. The IR was varied from 1.6 to 3.2 b/sbl/pol in steps of 0.1 b/sbl/pol. Figure 6.20 shows the NGMI and the post-decoding BER after 10,200 km of transmission. Post-decoding error-free transmission was achieved up to 2.8 b/sbl/pol. In addition, Fig. 6.21 shows constellations of the signal after the adaptive ML filters for several IRs up to 2.8 b/sbl/pol.

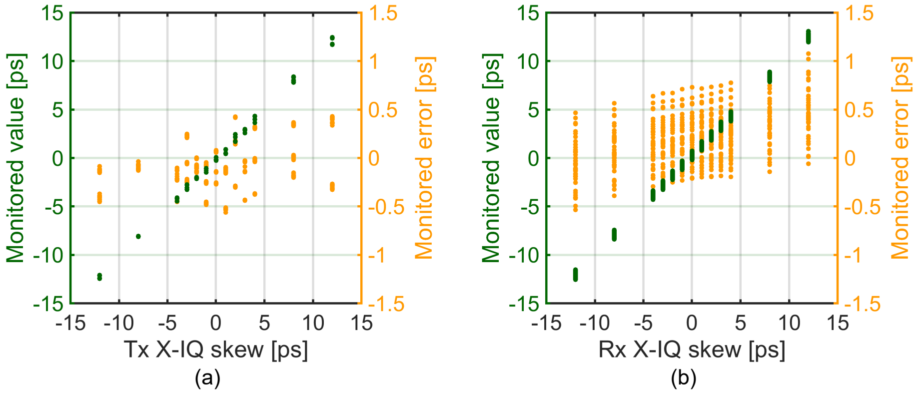


Figure 6.19. Impairment monitoring via the coefficients of the ML configuration after 10,200 km of SMF transmission for (a) Tx and (b) Rx IQ skew in the X polarization.

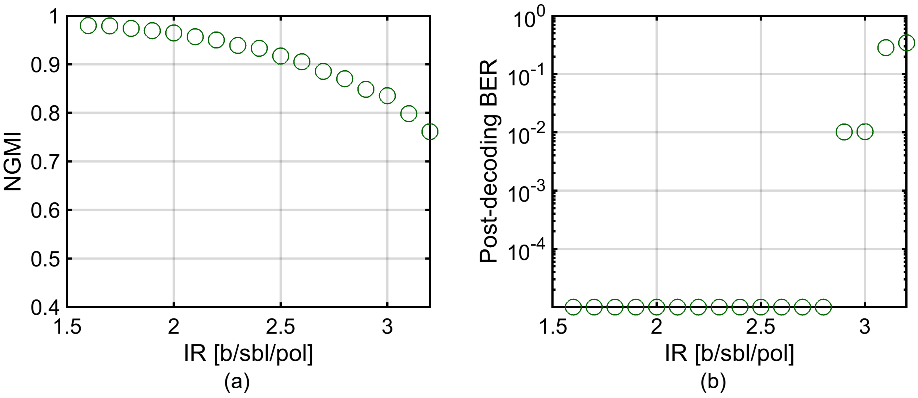


Figure 6.20. (a) NGMI and (b) post-decoding BER for reception of a 32-Gbaud PDM-64QAM signal by the ML configuration after 10,200 km of SMF transmission while varying the IR with fixed FEC.

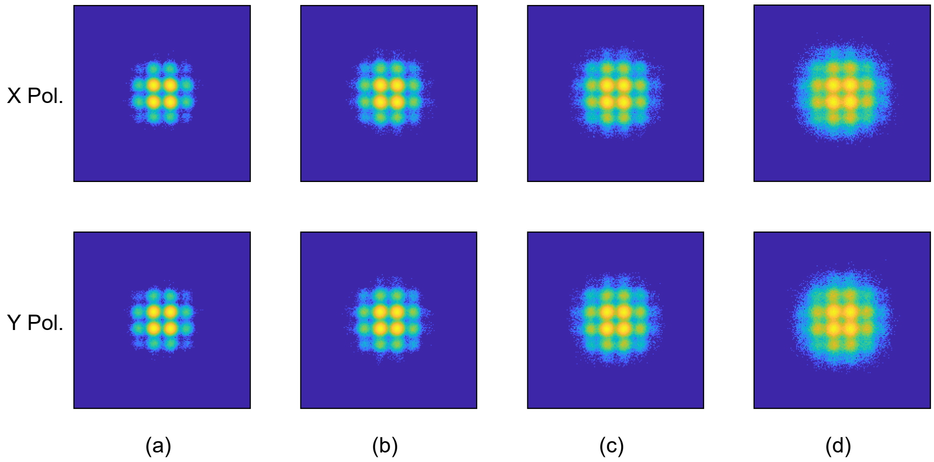


Figure 6.21. Constellations of a 32-Gbaud PDM-64QAM signal with IRs of (a) 1.6, (b) 2.0, (c) 2.4, and (d) 2.8 b/sbl/pol, as received by the ML configuration after 10,200 km of SMF transmission.

6.5 Summary

We have proposed an adaptive multi-layer filter architecture for Tx and Rx impairment compensation in ultra-long-haul transmission. In this architecture, CD compensation is performed on a received signal and its complex conjugate. The two signals are then given as augmented inputs to 2×1 SL MISO filters in the first layer of the proposed multi-layer filters, which is for Rx impairment compensation. The coefficients of the multi-layer filters are controlled by gradient calculation with back propagation from the last layer and SGD. The CD compensation filters are swept before the multi-layer filters, while the mutual non-commutativity of Rx IQ impairments and CD compensation is solved. This considerably mitigates the computational complexity of the back propagation in the case with the large CD accumulation. The proposed architecture leverages the features of adaptive multi-layer filters: namely, the filters in each layer can be individually designed, and impairments can be individually and simultaneously monitored from the converged coefficients.

We evaluated the proposed adaptive multi-layer filters with augmented inputs through both numerical simulation and in WDM transmission experiments. The experimental evaluation used 16 channels of 32-Gbaud PDM-PCS-64QAM signal with an IR of 2.4 b/sbl/pol in a 50-GHz grid over 10,200 km of SMF transmission, in which Tx and Rx IQ skews were digitally emulated. The results showed that the adaptive mult-layer filter architecture could compensate Tx IQ skew up to ±12 ps and Rx IQ skew over ± 20 ps under a large CD accumulation over the 10,000 km of SMF. Furthermore, emulated Tx and Rx IQ skews up to ±12 ps could be monitored individually from the adaptive filter coefficients of the corresponding layers with a maximum error of 1.1 ps.

Chapter 7

Frequency-domain adaptive MIMO filter with fractional oversampling using stochastic gradient descent and back propagation

Optimization by gradient descent and back propagation is widely applicable to embedded parameters in differentiable cascaded multiple layers, not only multi-layer FIR filters. In this chapter, we investigate a frequency-domain adaptive filter with fractional oversampling controlled by SGD and back propagation. This is an example of adaptive control of a filter that includes an internal sampling rate conversion. The sampling rate conversion can be treated as a layer in a multi-layer structure, and we can derive forward and back propagation for it.

We apply the fractional oversampling frequency-domain adaptive filter to mode demultiplexing in space-division multiplexed (SDM) transmission systems with coupled spatial channels to mitigate computational complexity. We evaluate the fractional oversampling frequency-domain adaptive MIMO filter in a long-haul transmission experiment over coupled 4-core fibers (C4CFs) with 16-channel WDM and 4-core SDM 32-Gbaud PDM-QPSK signals. Our findings show that the fractional oversampling frequency-domain adaptive 8×8 filter with $9/8\times$ oversampling utilized for mode demultiplexing provided little performance penalty after 6240-km transmission compared to the conventional $2\times$ oversampling frequency-domain adaptive 8×8 filter. The computational complexity in terms of the required number of complex-valued multiplications was reduced by 40.7%. This approach provides an efficient implementation of a MIMO filter for coupled SDM systems and also extends the application of adaptive filters to cases including a rate conversion.

7.1 Background

Capacity expansion of long-haul transmission over a conventional SMF is rapidly approaching a limitation [116,117]. The use of SDM transmissions that utilize a spatial degree of freedom in an optical fiber for signal multiplexing is one potential approach to overcoming this capacity crunch with SMFs [118,119]. In particular, SDM transmission systems with coupled spatial channels in which spatial coupling is allowed in a transmission line and compensated by MIMO DSP enable a high spatial density of channels within a limited cross-section of an optical fiber [115,120–122]. Coupled multi-core fibers (MCFs) are prime candidates for the transmission fibers of coupled SDM systems and exhibit favorable features for long-haul transmission, including low propagation loss and low spatial mode dispersion (SMD) thanks to the core design and arrangement [123]. Long-haul transmissions over coupled MCFs have been experimentally demonstrated along with mitigation of impairments over nonlinear fiber propagation by spatial coupling [124–126] and with real-time MIMO DSP implemented in field-programmable gate arrays [127].

In coupled SDM systems, the signal coupling among spatial modes and the temporal spread due to SMD that occur in a transmission line do not bind the transmission performance if a sufficiently long MIMO filter is used to compensate for them, though the number of spatial modes and SMD determine the scale of the MIMO filter and the required temporal length, respectively. The condition of spatial coupling over a coupled MCF temporally varies, as in the case of the polarization state in an SMF [128], and the SMD of current coupled MCFs is relatively large compared to polarization mode dispersion [123]. Therefore, a MIMO filter for a coupled SDM system is usually a large-scale adaptive filter with a large temporal spread, which requires high computational complexity. For this reason, an efficient MIMO filter implementation is crucial for the practical implementation of coupled SDM systems.

Regarding the scale of the MIMO filter, MIMO filter partitioning has been utilized to reduce complexity when a physical channel has low coupling between particular spatial modes [129,130], though it cannot be straightforwardly applied to a coupled MCF case with fully random coupling among all spatial modes. Thus, efficient implementation of a large temporal spread for the MIMO filter is a primary focus for SDM systems with fully random coupling. One approach is to use sub-carrier multiplexing [131], which was adopted in a real-time MIMO DSP demonstration [127]. Since the low symbol rate of one sub-carrier makes the symbol duration longer, designing the MIMO filter for one sub-carrier can have fewer taps to compensate for the same SMD without sub-carrier multiplexing. However, this approach does not directly reduce the total computational complexity for all the sub-carriers. Another approach is the use of a frequency-domain implementation with FFT and IFFT [132], which enables efficient implementation of a filter with a large temporal spread. Frequency-domain filters are often utilized for CD compensation that has a large temporal spread in long-haul SMF transmissions [16]. Adaptive frequency-domain filters have also been investigated for polarization demultiplexing in SMF transmission [133], as well as for compensation of the spatial coupling and SMD in coupled SDM systems [69,131,134].

Regardless of the time-domain or frequency-domain implementation, adaptive MIMO filters for coupled SDM systems have usually been operated with the signals at two-fold (2×) oversampling to obtain a fine equalization performance [127, 135]. In contrast, in SMF transmission systems, adaptive MIMO filters with input signals at a lower sampling rate with a non-integer oversampling factor of less than two (e.g., 1.5× oversampling, which is referred to as fractional oversampling here) have been investigated [136–140], as the symbol rate increases and a signal having a flat spectrum with a small roll-off factor is adopted. As in the case of sub-carrier multiplexing, decreasing the sampling rate of input signals to a filter lengthens the temporal duration per sample and enables a lower computational complexity. Fractionally spaced time-domain adaptive MIMO filters for SMF transmission systems have been implemented by using a polyphase filter to convert sampling rates before an adaptive filter [136,137] or implemented as a polyphase adaptive filter [138]. For frequency-domain implementation, a fractional oversampling filter has been investigated with channel estimation and sampling rate conversion [139]. For fully adaptive control of a fractional oversampling frequency-domain MIMO filter, the configuration in which the fractional oversampling frequency-domain inputs are expanded to those corresponding to 2× oversampling and the adaptive control of the filter coefficients is operated in the 2× oversampling domain [140]. However, these fractional oversampling adaptive MIMO filters have been investigated for transmission systems with an SMF, in which modal coupling and dispersion are relatively simple compared to the case with a coupled MCF. Moreover, the fractional oversampling adaptive MIMO filter in the frequency domain is based on the filter coefficient control in the 2× oversampling domain, not in the low fractional oversampling domain of the input signals, which leads to inefficiency in terms of the computational complexity.

7.2 Fractional oversampling frequency-domain adaptive MIMO filter controlled by SGD

In light of the above background, we investigate a fractional oversampling frequency-domain adaptive MIMO filter architecture in which the input signals are at fractional oversampling below 2× oversampling for long-haul transmission over a coupled MCF. Since we can usually have a criterion of error only for signals at the symbol rate, i.e., 1× sampling, it is difficult to use the direct output signals of the frequency-domain adaptive MIMO filter at fractional oversampling for adaptive coefficient control. Therefore, the frequency-domain sampling rate conversion to 1× sampling is placed after the fractional oversampling frequency-domain MIMO filter. The adaptive control of the filter coefficients is performed by SGD to minimize the loss function constructed with the output signals at 1× sampling in the time domain. Gradients are calculated with back propagation from the output signals, similar to the previous chapters regarding multi-layer filters. We derive the back propagation of the frequency-domain sampling rate conversion and show that it requires no complex-valued multiplications, which is an indicator of the computational complexity of DSP.

7.2.1 Frequency-domain adaptive filter from viewpoint of back propagation

We first review a conventional 2× oversampling frequency domain adaptive filter using the overlap-save method from the viewpoint of back propagation. Adaptive filter coefficient control of the conventional frequency domain adaptive filter [40,132] can be comprehended by gradient descent and back propagation by decomposing the frequency domain adaptive filter to multiple functional blocks. Forward and back propagation of the functional blocks appears in those of the fractional oversampling frequency-domain adaptive MIMO filter.

A schematic diagram of the conventional 2× oversampling frequency-domain adaptive filter with the overlap-save method is shown in Fig. 7.1. This diagram includes carrier recovery after down-sampling of the time-domain output signal to 1× sampling in order to adopt a symbol-error-based algorithm such as the LMS algorithm for the coefficient control [115]. The conventional 2× oversampling frequency-domain adaptive filter can also be implemented by utilizing an extended MIMO filter and divided input signals with even and odd indices [69,133][1], as shown in Fig. 7.2. We review here the model of the conventional 2× oversampling frequency-domain adaptive filter with the overlap-save method from the viewpoint of back propagation. We focus on the conventional architecture shown in Fig. 7.1 and consider the case with a single-input single-output filter for simplicity. The extension to a MIMO filter including the extended MIMO filter with the even-odd type shown in Fig. 7.2 is straightforward.

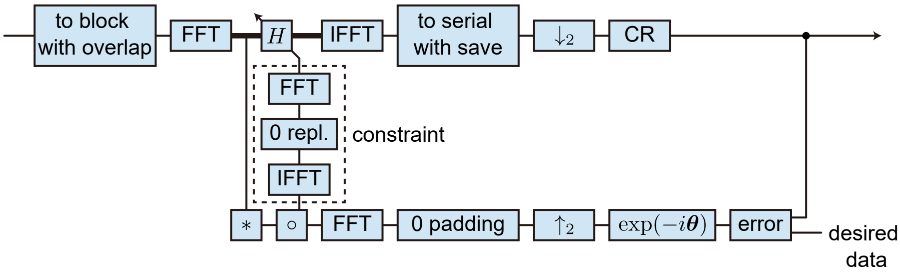


Figure 7.1. Schematic diagram of conventional 2× oversampling frequency-domain adaptive filter with overlap-save method. The bold line across H indicates a MIMO process.

Forward propagation

We consider the input and output signal blocks of the conventional 2× oversampling frequency-domain adaptive filter at a timing integer l given as

$$\boldsymbol{x} = (x[l], x[l-1], \ldots, x[l-N_x+1])^{\mathrm{T}}, \tag{7.1}$$

---

[1] This architecture is based on the noble identity of multi-rate systems [141].

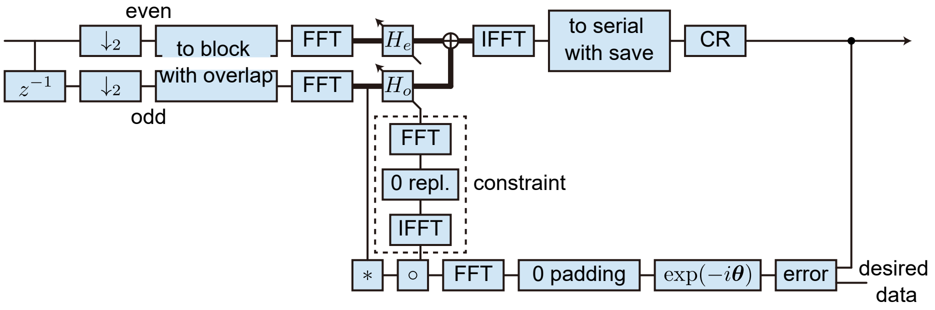


Figure 7.2. Schematic diagram of even-odd 2× oversampling frequency-domain adaptive filter with overlap-save method. The bold lines across H_e and H_o indicate MIMO processes.

$$\boldsymbol{y} = (y[l], y[l-1], \ldots, y[l-N_y+1])^{\mathrm{T}}, \tag{7.2}$$

where N_x and N_y are the lengths of the input and output signal blocks. The time-domain finite impulse response of the filter is given as

$$\boldsymbol{h} = (h[0], h[1], \ldots, h[M-1])^{\mathrm{T}}, \tag{7.3}$$

where M is the length of the finite impulse response. The input and output signal blocks are related by convolution as

$$\begin{pmatrix} y[l] \\ \vdots \\ y[l-N_y+1] \end{pmatrix} = \begin{pmatrix} h[0] & \cdots & h[M-1] & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & & & & & 0 \\ 0 & \cdots & 0 & h[0] & \cdots & h[M-1] \end{pmatrix} \begin{pmatrix} x[l] \\ \vdots \\ x[l-N_x+1] \end{pmatrix}, \tag{7.4}$$

where $N_y = N_x - M + 1$ as a result of convolution. Equation (7.4) is rewritten as

$$\boldsymbol{y} = H_{-}\boldsymbol{x}, \tag{7.5}$$

with a non-square matrix H_{-}. Expanding Eq. (7.4) so that the matrix become circulant, we obtain

$$\begin{pmatrix} y[l] \\ \vdots \\ y[l-N_y+1] \\ \tilde{y}[l-N_y] \\ \vdots \\ \tilde{y}[l-N_x+1] \end{pmatrix} = \begin{pmatrix} h[0] & \cdots & & h[M-1] & 0 & \cdots & 0 \\ 0 & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & & & & & & 0 \\ 0 & \cdots & 0 & h[0] & \cdots & & h[M-1] \\ h[M-1] & 0 & \cdots & 0 & h[0] & \cdots & h[M-2] \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ h[1] & \cdots & h[M-1] & 0 & \cdots & 0 & h[0] \end{pmatrix} \boldsymbol{x}, \tag{7.6}$$

which can be rewritten by using a circulant matrix H_+ as

$$\boldsymbol{y}_+ = \begin{pmatrix} \boldsymbol{y} \\ \tilde{\boldsymbol{y}} \end{pmatrix} = H_+ \boldsymbol{x}. \tag{7.7}$$

Since H_+ is circulant, it can be diagonalized with the discrete Fourier transform (DFT) matrix D as

$$H_+ = D^\dagger \Lambda D, \tag{7.8}$$

where † is a Hermitian conjugate. Λ has diagonal elements of $H = (H[0], \dots, H[N_x - 1])^\mathrm{T}$, which are related to the time-domain finite impulse response of the filler as

$$H = D\boldsymbol{h}_+, \tag{7.9}$$

$$\boldsymbol{h}_+ = \begin{pmatrix} \boldsymbol{h} \\ \boldsymbol{0} \end{pmatrix} = (h[0], \dots, h[M-1], 0, \dots, 0)^\mathrm{T}, \tag{7.10}$$

whose length is N_x. Thus, we obtain

$$\boldsymbol{y}_+ = D^\dagger \Lambda D \boldsymbol{x}, \tag{7.11}$$

and then

$$D\boldsymbol{y}_+ = \Lambda D \boldsymbol{x}, \tag{7.12}$$

where $D^{-1} = D^\dagger$. Multiplexing the DFT matrix D to a signal block corresponds to the conversion to the frequency-domain. We denote the frequency-domain representations of $\boldsymbol{y}_+$ and $\boldsymbol{x}$ as Y_+ and X. Finally, we obtain the model of the frequency-domain filter of the finite impulse response of $\boldsymbol{h}$ as

$$Y_+ = \Lambda X = H \circ X, \tag{7.13}$$

where H is the frequency-domain representation of $\boldsymbol{h}_+ = (\boldsymbol{h}\ \boldsymbol{0})^\mathrm{T}$. The output of the frequency-domain filter Y_+ is converted to $\boldsymbol{y}_+$ by IFFT, or by multiplexing D^{-1}. However, a portion of $\boldsymbol{y}_+$, $\tilde{\boldsymbol{y}}$, was added artificially to obtain Eq. (7.6), and it includes distortion due to wraparound from both ends of the input signal block. Thus, only $\boldsymbol{y}$ is saved from $\boldsymbol{y}_+$, and $\tilde{\boldsymbol{y}}$ is discarded. As a consequence, the time-domain output signal block after discard becomes shorter than that of the input signal block, and thus the input signal block is constructed with overlap between two successive blocks. These operations correspond to the overlap-save procedure [40]. According to the above model, the temporal spread of the time-domain response that can be represented with a frequency-domain filter without distortion due to wraparound corresponds to that of $\tilde{\boldsymbol{y}}$, that is, the overlap size of input signal blocks [40].

The time-domain output signal $\boldsymbol{y}$ after discard is at 2× oversampling. For symbol decision and calculating the symbol error for the adaptive coefficient update, 2× down-sampling is performed on $\boldsymbol{y}$ and we obtain the 1× sampling time-domain signal $\boldsymbol{y}_{\downarrow 2}$. The phase rotation of $\boldsymbol{\theta}$ is performed to $\boldsymbol{y}_{\downarrow 2}$ for CR, and the final output signal block of the conventional 2× oversampling frequency-domain adaptive filter with the overlap-save method is obtained as

$$\boldsymbol{z} = \boldsymbol{y}_{\downarrow 2} \circ \exp(i\boldsymbol{\theta}). \tag{7.14}$$

The computational graph of the process described above is shown in Fig. 7.3.

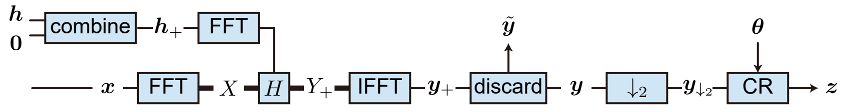


Figure 7.3. Computational graph of conventional 2× oversampling frequency-domain adaptive filter with overlap-save method for calculation of filter coefficients update with gradient descent and back propagation.

Back propagation

We then consider the adaptive control of the frequency-domain filter coefficients H. Since H is the frequency-domain representation of $\boldsymbol{h}_+$ and a portion of $\boldsymbol{h}_+$ is restricted to zero, we consider the update of the time-domain finite impulse response $\boldsymbol{h}$ by SGD and gradient calculation with back propagation according to the graph shown in Fig. 7.3, followed by the update of the frequency-domain filter coefficients H. Using SGD and Wirtinger derivatives as a basis, a coefficient ξ is updated to minimize the instantaneous loss function ϕ, as

$$\xi \to \xi - 2\alpha \frac{\partial \phi}{\partial \xi^*}, \tag{7.15}$$

where α is a step size. The superscript $*$ is complex conjugate. The loss function to be minimized should be real-valued, and thus the relation

$$\frac{\partial \phi}{\partial \xi^*} = \left(\frac{\partial \phi}{\partial \xi} \right)^*, \tag{7.16}$$

holds. The loss function consists of the final time-domain output signal block $\boldsymbol{z}$. When we use the LMS algorithm, the loss function is a norm of the difference of the signals and the desired symbols $\boldsymbol{d}$, as

$$\phi = ||\boldsymbol{d} - \boldsymbol{z}||^2, \tag{7.17}$$

where $||\ ||$ is the l_2 norm. The gradients of the loss function ϕ in terms of the final output signals $\boldsymbol{z}$ are calculated as

$$\frac{\partial \phi}{\partial \boldsymbol{z}} = -\boldsymbol{e}^*, \tag{7.18}$$

where $\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{z}$. According to the computational graph shown in Fig. 7.3, the gradients in terms of each variable are calculated with back propagation successively. The gradients in terms of $\boldsymbol{y}_{\downarrow 2}$ are

$$\frac{\partial \phi}{\partial \boldsymbol{y}_{\downarrow 2}} = \frac{\partial \phi}{\partial \boldsymbol{z}} \circ \exp(i\boldsymbol{\theta}). \tag{7.19}$$

The gradients in terms of $\boldsymbol{y}$ are obtained by 2× up-sampling with inserting zero, as

$$\frac{\partial \phi}{\partial \boldsymbol{y}} = \left(\frac{\partial \phi}{\partial \boldsymbol{y}_{\downarrow 2}} \right)_{\uparrow 2}. \tag{7.20}$$

The gradients in terms of $\boldsymbol{y}_+$ are obtained by inserting a zero vector that has the same size as the discarded $\tilde{\boldsymbol{y}}$, as

$$\frac{\partial \phi}{\partial \boldsymbol{y}_+} = \begin{pmatrix} \frac{\partial \phi}{\partial \boldsymbol{y}} \\ \mathbf{0} \end{pmatrix}. \tag{7.21}$$

The gradients in terms of the output signal block of the frequency-domain filter Y_+ are calculated by using the DFT matrix D, as

$$\frac{\partial \phi}{\partial Y_+} = D^{-1} \frac{\partial \phi}{\partial \boldsymbol{y}_+}. \tag{7.22}$$

The gradients in terms of the frequency-domain filter H are

$$\frac{\partial \phi}{\partial H} = \frac{\partial \phi}{\partial Y_+} \circ X, \tag{7.23}$$

and those in terms of the time-domain representation $\boldsymbol{h}_+$ are

$$\frac{\partial \phi}{\partial \boldsymbol{h}_+} = D \frac{\partial \phi}{\partial H}. \tag{7.24}$$

Since $\boldsymbol{h}_+ = (\boldsymbol{h}\ \mathbf{0})^{\mathrm{T}}$, the gradients in terms of the time-domain finite impulse response $\boldsymbol{h}$ are included in $\partial \phi / \partial \boldsymbol{h}_+$, as

$$\frac{\partial \phi}{\partial \boldsymbol{h}_+} = \begin{pmatrix} \frac{\partial \phi}{\partial \boldsymbol{h}} \\ \frac{\partial \phi}{\partial \tilde{\boldsymbol{h}}} \end{pmatrix}. \tag{7.25}$$

When the gradients of the loss function ϕ in terms of the time-domain finite impulse response $\boldsymbol{h}$ are given, $\boldsymbol{h}$ is updated by SGD to minimize the instantaneous loss function as

$$\boldsymbol{h} \to \boldsymbol{h} - 2\alpha \frac{\partial \phi}{\partial \boldsymbol{h}^*}. \tag{7.26}$$

Since $\boldsymbol{h}$ is related to the frequency-domain filter coefficients H as Eq. (7.9), Eq. (7.26) leads the update of the frequency-domain filter coefficients as

$$H \to H - 2\alpha D \begin{pmatrix} \frac{\partial \phi}{\partial \boldsymbol{h}^*} \\ \mathbf{0} \end{pmatrix}. \tag{7.27}$$

This equation is rewritten by using the gradients in terms of H, as

$$H \to H - 2\alpha D \left(\boldsymbol{c} \circ \left(D^{-1} \frac{\partial \phi}{\partial H^*} \right) \right), \tag{7.28}$$

where the elements of $\boldsymbol{c}$ are $c[i] = 1$ $(i \in [0, \ldots, M-1])$, $c[i] = 0$ $(i \in [M, \ldots, N_x - 1])$. The operation carried out on $\partial \phi / \partial H^*$ in the above equation corresponds to the constraint operation in the adaptive update of the frequency-domain filter coefficient [40, 69].

The above descriptions are how the model of the conventional 2× oversampling frequency-domain adaptive filter with the overlap-save method is constructed from the viewpoint of back propagation.

7.2.2 Fractional oversampling frequency-domain adaptive MIMO filter

A fractional oversampling frequency-domain adaptive MIMO filter architecture consists of a frequency-domain MIMO filter in which the input signals are processed at fractional oversampling rate and a frequency-domain sampling rate conversion to 1× sampling. The coefficients of the frequency-domain MIMO filter are controlled by SGD to minimize the loss function calculated from the time-domain outputs after the sampling rate conversion to 1× sampling. The gradients of the loss function in terms of the filter coefficients are calculated with back propagation from the outputs successively through the sampling rate conversion.

Figure 7.4 shows a schematic diagram of the fractional oversampling frequency-domain adaptive MIMO filter architecture. Here, the input signal's sampling rate is $M/L\times$ oversampling, where M and L are integers and we assume that $M/L \leq 2$. The fractional oversampling frequency-domain adaptive MIMO filter operates in the $M/L\times$ oversampling domain. The overlap-save method is used for frequency-domain implementation of a filter [40]. We describe here the case with a single-input single-output filter for simplicity, though the extension to a MIMO filter is straightforward. After the fractional $M/L\times$ oversampling frequency-domain MIMO filter, the sampling rate conversion to the 1× sampling domain is performed. The frequency-domain signal in the 1× sampling domain is converted to the time-domain with the overlap-save and then carrier recovery is executed.

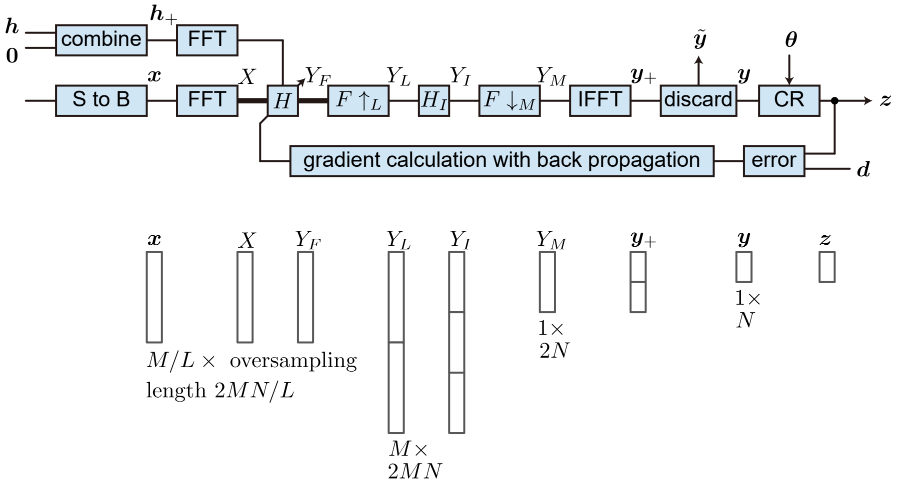


Figure 7.4. Schematic diagram of fractional oversampling frequency-domain adaptive filter with overlap-save method. The input signal x is at $M/L\times$ oversampling and the output signal z is at 1× sampling. The bold line across H indicates a MIMO process. The sizes of each signal block in the process are shown beneath, in which $L = 2$, $M = 3$, and the overlap is 50%. S to B: serial to block, FFT: fast Fourier transform, IFFT: inverse FFT.

Forward propagation

We first describe the forward propagation of this fractional oversampling frequency-domain adaptive MIMO filter architecture including the frequency-domain rate conversion to the 1× sampling domain. A frequency-domain filter deals with signals divided into blocks by the overlap-save method. We focus an input signal block at a timing integer l and treat complex-valued signals assuming DSP after coherent reception. The input signal block $\boldsymbol{x}$ at $M/L\times$ oversampling, that is, at the sampling rate of M/LR_s, is described as

$$\boldsymbol{x} = (x(lT_x), x((l-1)T_x), \ldots, x((l-N_x+1)T_x))^{\mathrm{T}}, \tag{7.29}$$

where the superscript T denotes the transpose. R_s is the symbol rate. The sampling interval is $T_x = LT/M$, where $T = 1/R_s$ is the symbol interval. The length of the input block $N_x = 2MN/L$ if the length of the output block at 1× sampling with 50% overlap/discard is N. The input signal block $\boldsymbol{x}$ is converted to its frequency-domain representation X by FFT. The length of X is the same as that of $\boldsymbol{x}$. The frequency-domain representation

$$X = (X(0), X(F_0), \ldots, X((N_x-1)F_0))^{\mathrm{T}}, \tag{7.30}$$

where $F_0 = MR_s/L/N_x = R_s/(2N)$, and the time-domain representation $\boldsymbol{x}$ are related by the Fourier transform as

$$X(kF_0) = \sum_{n=0}^{N_x-1} x\left((l-n)T_x\right) \exp\left(-i\frac{2\pi kn}{N_x}\right), \tag{7.31}$$

with an integer k, which can be rewritten as

$$X = D\boldsymbol{x}, \tag{7.32}$$

by using the DFT matrix D. The coefficients of the frequency-domain filter operating in the $M/L\times$ oversampling domain are represented by

$$H = (H(0), H(F_0), \ldots, H((N_x-1)F_0))^{\mathrm{T}}, \tag{7.33}$$

whose length is the same as N_x. H is the frequency-domain representation of $\boldsymbol{h}_+ = (\boldsymbol{h}\ \boldsymbol{0})^{\mathrm{T}}$. The direct output block Y_F of the frequency-domain filter operating in the $M/L\times$ oversampling domain is

$$Y_F = H \circ X. \tag{7.34}$$

The sampling rate conversion from the $M/L\times$ oversampling to 1× sampling domain is performed on Y_F. The sampling rate conversion consists of L-times up-sampling in the frequency-domain ($F\uparrow_L$), a decimation filter (H_I), and M-times down-sampling ($F\downarrow_M$). Y_F is first L-times up-sampled in the frequency-domain to Y_L. The time-domain representations of Y_F and Y_L are

$$\boldsymbol{y}_F = (y_F(lT_x), y_F((l-1)T_x), \ldots, y_F((l-N_x+1)T_x))^{\mathrm{T}}, \tag{7.35}$$

$$\boldsymbol{y}_L = (y_L(lLT_x/L), y_L((lL-1)T_x/L), \ldots, y_L((lL-LN_x+1)T_x/L))^{\mathrm{T}}, \tag{7.36}$$

respectively. The lengths of Y_L and $\boldsymbol{y}_L$ are both LN_x. According to L-times up-sampling by zero insertion, $\boldsymbol{y}_F$ and $\boldsymbol{y}_L$ are related as

$$y_L\left((lL-n)\frac{T_x}{L}\right)=\begin{cases}y_F((l-m)T_x) & (n=mL)\\ 0 & (\text{otherwise})\end{cases}, \tag{7.37}$$

where n, m are integers. Thus, the frequency-domain representations Y_F and Y_L are related as [141]

$$Y_L(kF_0)=Y_F(kF_0), \tag{7.38}$$

which is rewritten as

$$Y_L(kF_0)=Y_F(\mathrm{mod}(k,N_x)F_0), \tag{7.39}$$

considering the periodicity of Y_F. The L-times up-sampled signal block Y_L is then filtered by the decimation filter H_I, as

$$Y_I(kF_0)=Y_L(kF_0)H_I(kF_0), \tag{7.40}$$

where the filter coefficients of the decimation filter are

$$H_I(kF_0)=\begin{cases}L & \left(|kF_0|<\frac{M}{2L}R_s\right)\\ 0 & (\text{otherwise})\end{cases}, \tag{7.41}$$

to preserve the bandwidth of the input signal block at $M/L\times$ oversampling. The output signal block of the decimation filter Y_I is M-times down-sampled in the frequency domain to obtain the signal block in the 1× sampling frequency-domain Y_M. According to M-times down-sampling, the following relation holds [141]:

$$Y_M(kF_0)=\frac{1}{M}\sum_{m=0}^{M-1}Y_I((k-mN_y)F_0), \tag{7.42}$$

where $N_y=LN_x/M=2N$ is the length of Y_M.

The output signal block in the 1× sampling frequency-domain Y_M is converted to the time-domain signal $\boldsymbol{y}_+$ by IFFT. A portion of $\boldsymbol{y}_+=(\boldsymbol{y}\ \tilde{\boldsymbol{y}})^{\mathrm{T}}$ includes distortion due to wraparound from both ends of the input signal block via the frequency-domain filter, and thus the distorted portion $\tilde{\boldsymbol{y}}$ is discarded. Only $\boldsymbol{y}$ is saved based on the overlap-save method. The length of $\boldsymbol{y}$ is N with 50% overlap. To adopt a LMS-based loss function for adaptive control of the filter coefficient in this study, carrier recovery is performed on $\boldsymbol{y}$. The final output block of the fractional oversampling frequency-domain adaptive filter in the 1× sampling time-domain is

$$\boldsymbol{z}=\boldsymbol{y}\circ\exp(i\boldsymbol{\theta}), \tag{7.43}$$

where $\boldsymbol{\theta}$ is the phases of carrier recovery and exp represents the element-wise exponential function here.

Back propagation of frequency-domain sampling rate conversion

The coefficients of the fractional $M/L\times$ oversampling frequency-domain filter H are controlled by SGD to minimize the loss function constructed from the 1× sampling time-domain output $\boldsymbol{z}$. The gradients of the loss function in terms of the filter coefficients are calculated back propagation from the output block through the sampling rate conversion. We derive back propagation of the frequency-domain sampling rate conversion through Y_F to Y_M. We can find back propagation of the other elements (such as FFT/IFFT) in the adaptive control of a conventional 2× oversampling frequency-domain filter. Hereinafter, a frequency component is described as $X(kF_0) = X[k]$ for readability.

When the gradients of the loss function of ϕ in terms of the 1× sampling frequency-domain signal Y_M are given as $\partial\phi/\partial Y_M$, the gradients of ϕ in terms of Y_I can be calculated as

$$\frac{\partial\phi}{\partial Y_I[k]} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\partial\phi}{\partial Y_M[k+mN]}, \tag{7.44}$$

according to the chain rule of the derivatives by using Eq. (7.42). Similarly, from $\partial\phi/\partial Y_I$, the gradients of ϕ in terms of Y_L are calculated as

$$\frac{\partial\phi}{\partial Y_L[k]} = \frac{\partial\phi}{\partial Y_I[k]}H_I[k], \tag{7.45}$$

by using Eq. (7.40). From $\partial\phi/\partial Y_L$, the gradients of ϕ in terms of Y_F are calculated as

$$\frac{\partial\phi}{\partial Y_F[k]} = \sum_{l=0}^{L-1}\frac{\partial\phi}{\partial Y_L[k+lN]}, \tag{7.46}$$

with Eq. (7.39). These are back propagations of the frequency-domain sampling rate conversion from the $M/L\times$ to 1× oversampling domain. Since we can calculate $\partial\phi/\partial Y_M$ from $\boldsymbol{z}$ and $\partial\phi/\partial H$ or $\partial\phi/\partial\boldsymbol{h}$ from $\partial\phi/\partial Y_F$ with back propagation, we now obtain an adaptive update for the filter coefficients H of the fractional $M/L\times$ oversampling frequency-domain filter by SGD.

7.2.3 Computational complexity

We evaluated the computational complexity of the fractional $M/L\times$ oversampling frequency-domain adaptive MIMO filter and compared it with that of the conventional 2× oversampling frequency-domain adaptive MIMO filter. The number of complex-valued multiplications is used as an indicator of the computational complexity of DSP. The temporal spread of a frequency-domain filter response with the overlap-save method without any distortion due to wraparound corresponds to the spread of the overlap region of the input signal blocks [40, 142], which is confirmed in the experimental evaluation described later. We compared the two filters under the same temporal spread, or filter length N, in units of symbol periods, which results in different filter lengths in units of sample periods when we compare filters in different oversampling domains.

In conventional frequency-domain filters, the input block length is usually set to a power of two, and the overlap ratio is set to 50% [16] so that the required FFT/IFFT sizes become powers of two, which enables the use of efficient FFT/IFFT algorithms such as Radix-2 [143]. In contrast, as described in the previous section, the fractional $M/L\times$ oversampling frequency-domain filter requires FFT/IFFT whose sizes are $N_x = 2MN/L$ and $N_y = 2N$ under the overlap ratio of 50%. These sizes are not restricted to powers of two with general M and L. By using appropriate algorithms such as the prime-factor FFT, we can perform FFT/IFFT efficiently even when n is not a power of two [143,144]. We consider the case with N of a power of two and also L of a small power of two. In this case, N_x becomes the multiplication of a power of two and a few small factors, which is suitable for applying the prime-factor FFT. We assume here that the required number of complex-valued multiplications per FFT/IFFT is given as $\mu(n) = n\log_2(n)/2$ [145], for any integer n.

The computational complexities of the conventional 2× oversampling frequency-domain adaptive MIMO filter, the even-odd 2× oversampling frequency-domain adaptive MIMO filter [69,133], and the fractional $M/L\times$ oversampling frequency-domain adaptive MIMO filter are estimated when the filter length is N symbols, the overlap ratio is 50%, and the scale of the MIMO filter is $K\times K$. As described previously, in the frequency-domain sampling rate conversion and its back propagation in the fractional $M/L\times$ oversampling frequency-domain adaptive MIMO filter, only multiplication by H_I appears, which corresponds to multiplication by a fixed value L or zero. Considering that multiplication by L in forward and back propagation can be accommodated in normalization of signals or step size adjustment, the frequency-domain sampling rate conversion and its back propagation do not contribute to the required complex-valued multiplications, while it requires more memory for the larger intermediate signals. Table 7.1 lists the required number of complex-valued multiplications per output symbol for forward propagation, back propagation for the coefficient update, and the constraint operation for the frequency-domain filter [40,69] for the three compared architectures by using $\mu(n)$. Forward propagation is always required for every input signal block to obtain output signals. Back propagation can be carried out intermittently if we regard the channel variation as slow. It has been reported that the constraint operation can be omitted for less complexity in some cases for an SDM transmission [69], but we assume here that back propagation and the constraint operation are carried out for every block.

Table 7.1: Required complex-valued multiplications per output symbol.

	conventional 2×	even-odd 2×	fractional $M/L\times$
Forward	$4K + 2\frac{\mu(4N)}{N} + 1$	$4K + 3\frac{\mu(2N)}{N} + 1$	$2K\frac{M}{L} + \frac{\mu(2MN/L)}{N} + \frac{\mu(2N)}{N} + 1$
Back	$4K + \frac{\mu(4N)}{N} + 1$	$4K + \frac{\mu(2N)}{N} + 1$	$2K\frac{M}{L} + \frac{\mu(2N)}{N} + 1$
Constraint	$2K\frac{\mu(4N)}{N}$	$4K\frac{\mu(2N)}{N}$	$2K\frac{\mu(2MN/L)}{N}$

Figure 7.5(a) shows the required number of complex-valued multiplications per output symbol while changing the filter length N for the conventional 2× oversampling frequency-domain adaptive MIMO filter, the even-odd 2× oversampling frequency-domain adaptive MIMO filter, the fractional $M/L\times$ oversampling frequency-domain adaptive MIMO filter, and the conventional 1× sampling frequency-domain adaptive MIMO filter. The two cases with $M = 3$, $L = 2$ and with $M = 9$, $L = 8$ are shown for the fractional $M/L\times$ oversampling frequency-domain adaptive MIMO filter. The number of spatial modes K was set to eight considering four-core and two-polarization transmission over C4CFs. We can see here that the fractional oversampling frequency-domain adaptive MIMO filter reduces the computational complexity as the oversampling ratio

M/L decreases. The fractional $3/2\times$ and $9/8\times$ oversampling frequency-domain adaptive MIMO filters show a lower computational complexity compared to the conventional and even-odd $2\times$ oversampling frequency-domain adaptive MIMO filters. At $N = 256$, which was used in the long-haul transmission experiment over C4CFs described later, the fractional $9/8\times$ oversampling frequency-domain adaptive MIMO filter reduces the required complex-valued multiplications by 48.1% compared to the conventional $2\times$ oversampling frequency-domain adaptive MIMO filter and by 40.7% compared to the even-odd $2\times$ oversampling frequency-domain adaptive MIMO filter. Figure 7.5(b) shows the required number of complex-valued multiplications per output symbol for the fractional $M/L\times$ oversampling frequency-domain adaptive MIMO filter when the filter length N is 256 symbols while changing the oversampling factor.

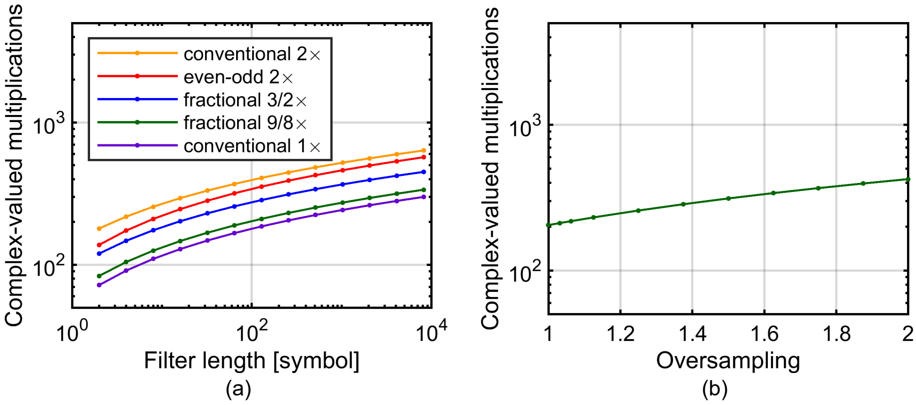


Figure 7.5. (a) Required number of complex-valued multiplications per output symbol of conventional $2\times$ oversampling frequency-domain adaptive MIMO filter, even-odd $2\times$ oversampling frequency-domain adaptive MIMO filter, fractional oversampling frequency-domain adaptive filter with $M/L = 3/2$ and $M/L = 9/8$, and conventional $1\times$ sampling frequency-domain adaptive MIMO filter. The overlap ratio and the number of modes K are 50% and 8, respectively. (b) Dependence of number of complex-valued multiplications per output symbol of the fractional oversampling frequency-domain adaptive filter on the oversampling factor with the filter length N of 256 symbols.

7.3 Evaluation in transmission experiment over coupled 4-core fibers

We evaluated the performance of the fractional oversampling frequency-domain adaptive MIMO filter to determine whether a low oversampling ratio with negligible degradation is feasible. In this evaluation, we applied the fractional oversampling frequency-domain adaptive MIMO filter for compensation of modal coupling and SMD in a long-haul transmission of 16-channel WDM and 4-core SDM 32-Gbaud PDM-QPSK signals over C4CFs.

Experimental setup

The experimental setup is shown in Fig. 7.6, where 16-channel WDM and 4-core SDM 32-Gbaud PDM-QPSK signals were transmitted over C4CFs in a recirculating loop configuration. The fractional oversampling frequency-domain adaptive MIMO filter was applied for mode demultiplexing, that is, compensation of modal coupling and dispersion, in the receiver-side DSP.

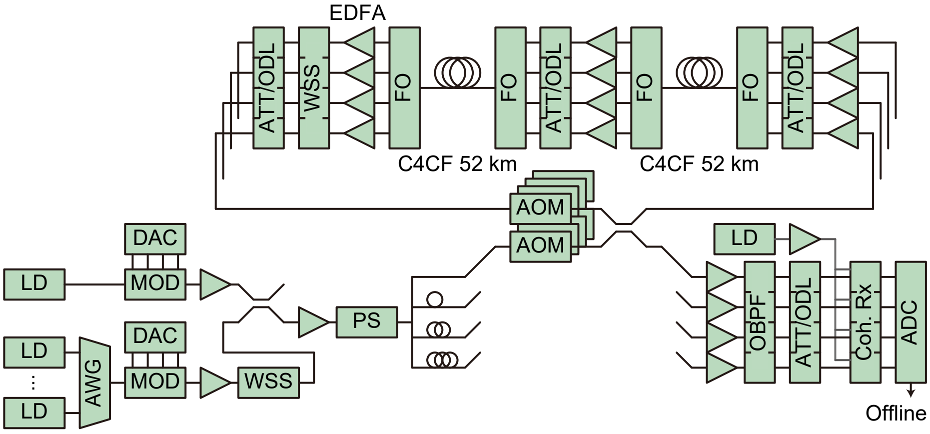


Figure 7.6. Experimental setup for WDM/SDM transmission of 32-Gbaud PDM-QPSK signals over a loop configuration of two spans of 52-km C4CF. ATT: optical attenuator, ODL: optical delay line, FO: fan-out.

On the transmitter side, 16-channel WDM signals of 32-Gbaud PDM-QPSK on a 50-GHz grid at frequencies of 192.90 to 193.65 THz were prepared. The signal at 193.3 THz was generated with a laser source having a linewidth of about 100 kHz modulated by electrical signals from a four-channel DAC at a sampling rate of 64 GS/s and vertical resolution of eight bits. The transmitted data consisted of a FEC frame of the low-density parity-check code for DVB-S2 with a frame length of 64,800 and a code rate of 0.8 for each IQ component and two polarizations, i.e., four frames in total. The data were mapped to PDM-QPSK and up-sampled to 2× oversampling so that they could be generated using the DAC. Root-raised cosine filtering with a roll-off factor of 0.1 and pre-compensation of the frequency characteristics of the transmitter devices were carried out on the signals. The remaining 15 channels were generated in a similar manner by using another modulator and 4-channel DAC with 15 channels of laser sources multiplexed by an arrayed waveguide grating. Different random bits were loaded in the FEC payload. The optical signal at 193.3 THz and the remaining 15 signals were amplified using EDFAs and combined after channel power equalization by a WSS. After low-speed polarization scrambling at $10\times2\pi$ rad/s, the WDM signals were split into four and each was sufficiently decorrelated (by 5 m of an SMF each) to emulate four-core SDM signals.

We evaluated the four-core SDM 32-Gbaud PDM-QPSK signals at 193.3 THz. Figure 7.7 shows the optical spectrum of the transmitted 32-Gbaud PDM-

QPSK signal measured with a high-resolution coherent optical spectrum analyzer. As we can see, the spectrum has an almost flat-top profile with the edges at ±16 GHz, reflecting the fact that a root-raised cosine filter with a roll-off factor of 0.1 was executed. The 1× and 2× oversampling for 32-Gbaud signals correspond to 32 and 64 GS/s, whose Nyquist frequencies are 16 and 32 GHz, respectively. The flat-top spectrum shown in Fig. 7.7 indicates that sampling near 1× is sufficient to reconstruct the signal from the sampled waveform.

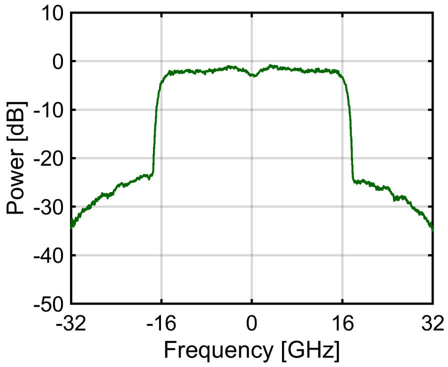


Figure 7.7. Optical spectrum of signal under evaluation measured by coherent optical spectrum analyzer with resolution of 150 MHz.

The transmission line in a recirculating loop configuration consists of switches with AOMs, two spans of 52-km C4CF with optical amplification by parallel EDFAs, and WSSs to compensate for gain/loss over WDM signals in a loop. The C4CFs had four cores arranged in a square with a core pitch of 20 μm. The clad diameter, averaged propagation loss, effective area of cores, and SMD were 125 μm, 0.165 dB/km, 112 $\mu\mathrm{m}^2$, and 6.9 ps/$\sqrt{\mathrm{km}}$, respectively. A 52-km C4CF was connected to SMF-based devices such as EDFA via fan-out (FO) devices that were fusion-spliced at both ends of the C4CF span. The relative delays in SMF-based parts in the loop were adjusted to be as small as possible by utilizing optical delay lines (ODLs). The span input optical power was set at the input of the FO at one end of a C4CF by using optical attenuators (ATTs). On the basis of a preliminary experiment with eight WDM channels, we set the span input optical power to 0 dBm/ch/core, which was slightly adjusted in consideration of the four cores to reduce the mode-dependent loss on the basis of the transmission performance. The averaged optical signal-to-noise ratio after two spans was 36.3 dB/0.1 nm.

The optical signals after transmission in the recirculating loop of the C4CFs were amplified by EDFAs, and the signals at 193.3 THz were demultiplexed by OBPFs having a 3-dB bandwidth of 50 GHz. The signals were then received by four polarization-diversity coherent receivers. A laser source having a linewidth of about 4 kHz was divided into four and supplied to coherent receivers as a local oscillator. The four-channel outputs of the four coherent receivers were sampled by a 16-channel oscilloscope used as an ADC at 50 GS/s and a vertical resolution of eight bits. Successive waveforms of 1MS were acquired five times for each transmission condition and processed by offline DSP.

A schematic diagram of the offline DSP is shown in Fig. 7.8. The received waveforms were first normalized and resampled to 2× oversampling on the basis of the timing error signal [86] after CD compensation for every core. CD compensation and matched filtering were carried out individually for eight signals of four cores and two polarizations. Frame synchronization was then performed to use the data-aided LMS algorithm for the adaptive coefficient update of a MIMO filter to avoid degenerated outputs [146] and ensure fine convergence. Because of the restriction in this experiment in which SDM signals were emulated by dividing a signal and decorrelation, we assume the transmitted data frames are a known training pattern at this stage. Frame header detection was carried out in a correlation-based manner after resampling to 1× sampling with a training pattern, though it does not work under a frequency offset between a carrier frequency and an LO [147]. Thus, a maximum peak of the correlation was searched for while the frequency offset compensation was performed and the amount of the frequency offset was swept in a step of 1 MHz. After detecting the frame header, frame synchronization and frequency offset compensation were carried out on the waveforms in the 2× oversampling domain.

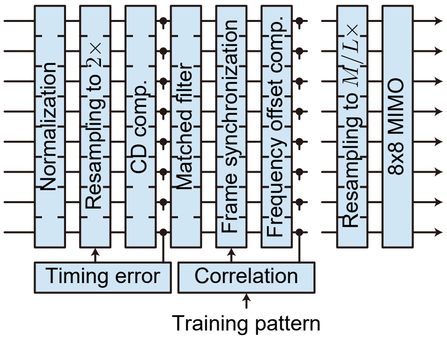


Figure 7.8. Offline DSP including proposed fractional oversampling frequency-domain adaptive 8×8 MIMO filter for mode demultiplexing.

The 2× oversampling waveforms after frame synchronization and frequency offset compensation were then resampled to $M/L\times$ oversampling for the fractional oversampling frequency-domain adaptive 8×8 MIMO filter shown in Fig. 7.4 for mode demultiplexing. Several oversampling ratios M/L below two were evaluated. The overlap ratio for the frequency-domain filter was set to 50%. The adaptive control of the filter coefficients was first carried out by the loss function based on the data-aided LMS algorithm. After convergence, the adaptive control was switched to that based on the decision-directed LMS. The step size was optimized for each oversampling ratio M/L[2]. Carrier recovery in the fractional oversampling frequency-domain adaptive filter was carried out us-

---

[2] We found that the optimal step sizes were similar when we set $H_I[k] = 1$ in Eq. (7.45), regardless of M/L.

ing a phase-locked loop with data-aided or decision-directed phase errors for eight spatial modes individually. After convergence with the decision-directed LMS algorithm, the Gram-Schmidt orthogonalization procedure was carried out on the time-domain 1× sampling output signals of the fractional oversampling frequency-domain adaptive 8×8 MIMO filter for compensation of a phase deviation between in-phase and quadrature components due to the imperfection of the bias control of the modulator in the transmitter [8]. Finally, the pre- and post-FEC BERs were evaluated. As a baseline, we also evaluated the DSP architecture with the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter. In this case, the 2× oversampling waveforms after frame synchronization and frequency offset compensation were fed into the frequency-domain adaptive 8×8 filter.

Note that we resampled the signals to $M/L\times$ oversampling just before the fractional oversampling frequency-domain adaptive 8×8 MIMO filter (as shown in Fig. 7.8) in order to focus specifically on its performance. This DSP architecture is indeed implementable, though the computational complexity of DSP can be further reduced if CD compensation and the matched filtering are carried out at M/L oversampling below 2× oversampling when the sampling rate of the ADC itself is M/L oversampling.

Dependence on oversampling ratio M/L

We first evaluated the performance of the fractional $M/L\times$ oversampling frequency-domain adaptive 8×8 MIMO filter while changing M and L at the transmission distance of 6240 km, or 60 loops. The filter length was set to 256 in units of symbol periods for any oversampling ratio M/L, which was optimal and sufficient to compensate for modal dispersion in this experiment, as shown later. L was set to a small power of two. Figure 7.9 shows the pre-FEC BERs averaged over eight spatial and polarization modes for five waveform acquisitions as a function of the oversampling ratio. The results with the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter are plotted at the point of the oversampling ratio of two in Fig. 7.9. We can see here that little performance penalty occurs down to the oversampling ratio of 9/8, or 1.125. According to Fig. 7.9, a slight degradation occurs at 17/16× oversampling, and the performance degradation increases drastically as the oversampling ratio decreases below one. As described in the previous section, the fractional 9/8× oversampling frequency-domain adaptive filter requires FFT and IFFT whose sizes are $3^2 \cdot 2^{-2}N$ and $2N$ under the overlap ratio of 50%. When N is a power of two, they consist of the multiplication of a power of two and small prime factors, resulting in a reasonable computational complexity, as shown in Fig. 7.5. At $N = 256$, the fractional 9/8× oversampling frequency-domain adaptive MIMO filter provides negligible degradation compared to the conventional 2× oversampling frequency-domain adaptive MIMO filter, with a 48.1% reduction of the complex-valued multiplications. In the following evaluations, we use $M = 9$ and $L = 8$ for the fractional oversampling frequency-domain adaptive 8×8 MIMO filter.

Figures 7.10 and 7.11 show the constellations of eight spatial modes after demodulation with the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter and after the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter, respectively. No significant difference between them

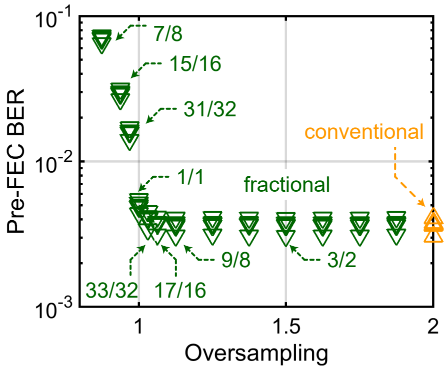


Figure 7.9. Dependence of pre-FEC BER with fractional oversampling frequency-domain adaptive 8×8 MIMO filter for reception at 6240 km on oversampling ratio of input signal. Pre-FEC BERs averaged over eight spatial modes for five waveform acquisitions are plotted at each oversampling ratio.

can be observed.

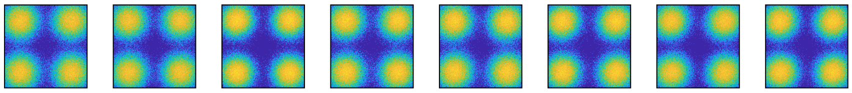


Figure 7.10. Constellations of eight spatial modes after demodulation with conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter for reception at 6240 km.

Dependence on filter length

We next evaluated the performance of the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter and the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter while changing the filter length N at the same transmission distance of 6240 km. Figure 7.12(a) shows the dependence of the pre-FEC BERs on the filter length in units of symbol periods. The pre-FEC BERs averaged over eight spatial modes for five waveform acquisitions are shown. The results with the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter and the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter show similar behavior. When the filter length was smaller than the temporal spread to be compensated due to modal dispersion in the transmission line, the performance was degraded due to inadequate compensation. In contrast, when the filter length was much larger than the

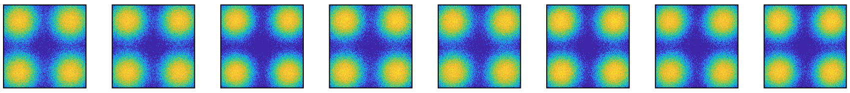
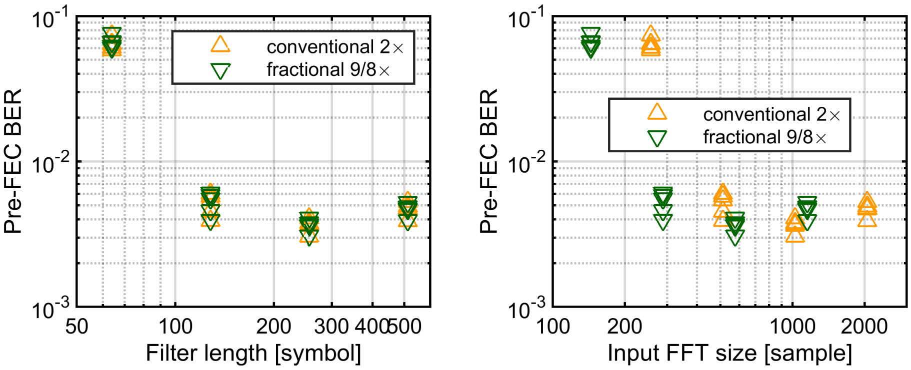


Figure 7.11. Constellations of eight spatial modes after demodulation with fractional-spaced 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter for reception at 6240 km.

temporal spread to be compensated, it caused excess errors of the filter coefficient due to the adaptive control by SGD with a finite step size [41]. The optimal filter length was 256 symbols for both filters in this experiment. Since the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter operated in the lower oversampling domain, the FFT size for the input signals decreased for the same filter length compared to the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter. Figure 7.12(b) shows the same results as a function of the input FFT size in units of sample periods. The fractional oversampling frequency-domain adaptive filter requires a smaller FFT size in units of sample periods.



(a) (b)

Figure 7.12. Dependence of pre-FEC BER on (a) the filter length in units of symbol period and (b) the input FFT size in number of samples. The results with the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter and with the fractional-spaced 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter for reception at 6240 km are plotted.

Convergence of loss

We compared the convergence of the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter and the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter at the transmission distance of 6240 km. Figure 7.13 shows an example of the convergence of the loss function by the data-aided LMS algorithm with both filters as a function of time in units of the symbol period. The filter length was set to 256 symbols. The adaptive filter coefficient was updated every output signal block, and thus both filters were updated every 256 symbols. The convergence behaviors of the filters were similar in terms of the time in units of symbol periods. Therefore, we can conclude that the two filters have comparable convergence properties.

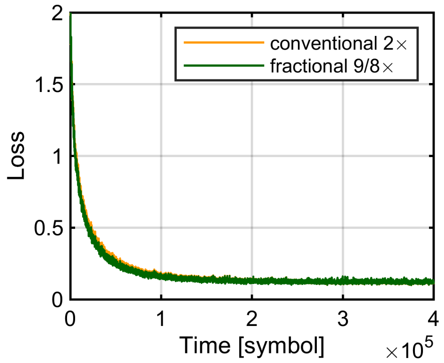


Figure 7.13. Convergence of loss with conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter and with fractional-spaced 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter for reception at 6240 km. The loss variations were moving-averaged over 100 symbols.

Transmission performance

Finally, we evaluated the transmission performances over C4CFs with the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter and the conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter. The filter length was again set to 256 symbols. Figure 7.14 shows the pre- and post-FEC BERs averaged over eight spatial modes for five waveform acquisitions as a function of the transmission distance in the case with the conventional filter. The results with no bit errors are plotted at the level of 10^{-6} for visibility. Figure 7.15 shows the pre- and post-FEC BERs in the case with the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter. In the both cases, all error-free transmissions after FEC for five acquisitions were achieved up to 7280 km. The transmission performances of the two were similar. Therefore, we can confirm that the fractional 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter is feasible for long-haul transmission over C4CFs.

7.4 Summary

We proposed a fractional oversampling frequency-domain adaptive MIMO filter architecture in which the input signals are at fractional oversampling below 2×

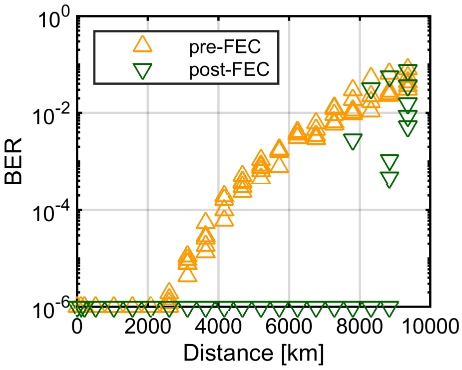


Figure 7.14. Pre- and post-FEC BER after long-haul transmission over C4CFs with conventional 2× oversampling frequency-domain adaptive 8×8 MIMO filter.

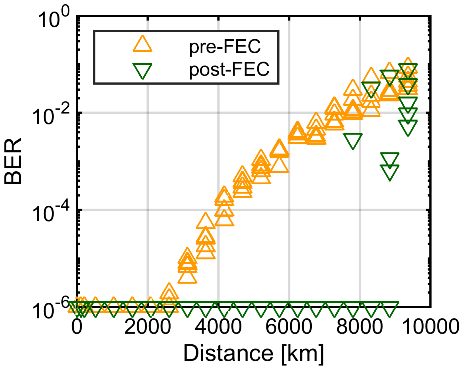


Figure 7.15. Pre- and post-FEC BER after long-haul transmission over C4CFs with fractional-spaced 9/8× oversampling frequency-domain adaptive 8×8 MIMO filter.

oversampling for long-haul transmission over a coupled MCF. The frequency-domain sampling rate conversion to 1× sampling is placed after the fractionally spaced frequency-domain MIMO filter. Stochastic gradient descent and gradient calculation with back propagation through the sampling rate conversion from the output signals are used for adaptive control of the filter coefficients to minimize the loss function constructed with the output signals at 1× sampling in the time domain. We evaluated the proposed filter in a long-haul transmission experiment over a C4CF with 16-channel WDM and 4-core SDM 32-Gbaud PDM-QPSK signals. The fractional oversampling frequency-domain adaptive 8×8 filter with 9/8× oversampling used for mode demultiplexing provided little performance penalty after 6240-km transmission compared to the conventional 2× oversampling frequency-domain adaptive 8×8 filter, and the computational complexity in terms of the required number of complex-valued multiplications was reduced by 40.7% compared to the even-odd 2× oversampling frequency-domain adaptive MIMO filter.

Chapter 8

Conclusion and outlook

Efficient and precise impairment compensation for optical fiber communication systems was investigated. By considering the mutual non-commutativity of impairments and compensation of them, the adaptive multi-layer filter architecture, whose filter coefficients were controlled by SGD and back propagation to minimize the loss function consisting of the last layer outputs, was proposed and evaluated.

The multi-layer filter architecture consisting of SL and WL filters to compensate for relevant impairments that occur in an optical fiber communication system including Tx and Rx impairments was proposed in Chapter 3. The adaptive filter coefficient control algorithm for multi-layer SL and WL FIR filters with SGD and back propagation was derived. The compensation capability of the adaptive multi-layer SL&WL filters for IQ skew in both Tx and Rx was evaluated through simulations with 32-Gbaud PDM-QPSK and an experiment on 32-Gbaud PDM-64QAM transmission over a 100-km SMF span. The results showed that they could compensate for IQ skew in both Tx and Rx in the presence of CD, polarization rotation, and a frequency offset.

The adaptive multi-layer filters compensate the relevant impairments in the corresponding layers. The monitoring method for individual impairments in a Tx and Rx including IQ skew, IQ imbalance, and IQ phase deviation on the basis of the filter coefficients of our adaptive multi-layer SL&WL filter architecture was proposed in Chapter 4. The impairment monitoring was evaluated through simulations using 32-Gbaud PDM-QPSK and a 100-km SMF transmission experiment of 32-Gbaud PDM-64QAM. The results indicated that IQ skew, IQ imbalance, and IQ phase deviation in both Tx and Rx could be individually monitored from the corresponding WL filter coefficients precisely and simultaneously when multiple impairments existed.

The adaptive multi-layer filter architecture was extended by incorporating nonlinear Volterra filters positioned to compensate for nonlinearity that occurs in Tx and Rx components under other impairments such as CD in Chapter 5. The performance of the adaptive multi-layer filters including Volterra filter layers numerically and experimentally in the transmission of a 23 Gbaud PDM-64QAM signal over one span of a 100-km SMF. The results demonstrated that the adaptive multi-layer filters including Volterra filter layers could compensate for the nonlinearity that occurs in both Tx and Rx simultaneously and effectively when other impairments coexisted.

In order to mitigate the computational complexity for back propagation of the multi-layer filters, the configuration, in which the CD compensation filters are swept before the multi-layer filters and the signal and its complex conjugate after CD compensation are the augmented inputs of the multi-layer filters, was proposed. The proposed adaptive ML filters with augmented inputs through both numerical simulation and inWDM transmission experiments. The results showed that this adaptive mult-layer filter architecture could compensate Tx and Rx IQ skew in the ultra-long-haul transmission over the 10,000 km of SMF. Furthermore, simultaneous and individual Tx and Rx IQ skew monitoring was also confirmed.

The frequency-domain adaptive filter with fractional oversampling based on SGD and back propagation of multiple functional layers was proposed for efficient mode demultiplexing in SDM transmission systems with coupled spatial channels in Chapter 7. Back propagation for the frequency-domain sampling rate conversion was derived. The fractional oversampling frequency-domain adaptive filter was evaluated in a long-haul transmission experiment over a C4CF with 16-channel WDM and 4-core SDM 32-Gbaud PDM-QPSK signals. The fractional oversampling frequency-domain adaptive 8×8 filter with 9/8× oversampling used for mode demultiplexing provided little performance penalty after 6240-km transmission compared to the conventional 2× oversampling frequency-domain adaptive 8×8 filter, with the reduced computational complexity.

In consequence, compensation of impairments for optical fiber communication systems, especially of Tx and Rx impairments, precisely and efficiently in terms of the computational complexity was achieved by the adaptive multi-layer filter architectures designed for a target. Tx and Rx impairments due to device imperfection are becoming more and more critical for optical fiber communication systems as they adopt more advanced modulation formats and higher symbol rates.

This study is also one example of the application of optimization with gradient descent and back propagation for cascaded multiple functional blocks, which can be found recently in longitudinal monitoring of an optical fiber communication systems based on optimization of some cascaded multi-blocks in the Rx DSP [148, 149], for example. It was shown that the adaptive control by SGD and back propagation works well by designing each layer in multi-layer filters by imposing several constraints based on the physical model of an optical fiber communication system. This study paves a way between the field of optical communication systems and that of machine learning.

We hope that this study contributes to future optical fiber communication systems and information processing.

Publication List

Related Journal Papers

1. Manabu Arikawa and Kazunori Hayashi, "Adaptive equalization of transmitter and receiver IQ skew by multi-layer linear and widely linear filters with deep unfolding," Opt. Express **28**(16), 23478 (2020)[1].
2. Manabu Arikawa and Kazunori Hayashi, "Transmitter and receiver impairment monitoring using adaptive multi-layer linear and widely linear filter coefficients controlled by stochastic gradient descent," Opt. Express **29**(8), 11548 (2021)[2].
3. Manabu Arikawa and Kazunori Hayashi, "Adaptive multi-layer filters incorporated with Volterra filters for impairment compensation including transmitter and receiver nonlinearity," Opt. Express **29**(18), 28366 (2021)[3].
4. Manabu Arikawa, Masaki Sato, and Kazunori Hayashi, "Compensation and monitoring of transmitter and receiver impairments in 10,000-km single-mode fiber transmission by adaptive multi-layer filters with augmented inputs," Opt. Express **30**(12), 20333 (2022)[4].
5. Manabu Arikawa, Mingqi Wu, Keisuke Yasuhara, Daishi Masuda, Kazunori Hayashi, Takanori Inoue, and Yoshihisa Inada, "Long-haul WDM/SDM transmission over coupled 4-core fibers installed in submarine cable," J. Light. Technol. **41**(6), 1649 (2023)[5].
6. Manabu Arikawa and Kazunori Hayashi, "Frequency-domain adaptive MIMO filter with fractional oversampling using stochastic gradient descent for long-haul transmission over coupled 4-core fibers," Opt. Express **31**(8), 13104 (2023)[6].

Related Peer-Reviewed Conference Proceedings

1. Manabu Arikawa and Kazunori Hayashi, "Adaptive multi-layer filters for compensating for impairments in transmitters and receivers for SDM trans-

---

[1] Chapter 3.

[2] Chapter 4.

[3] Chapter 5.

[4] Chapter 6.

[5] The concept described in Chapter 6 was applied to SDM transmission in which more IQ channels are used in the Tx and Rx.

[6] Chapter 7.

mission," ECOC 2022, Tu5.25.

Other Journal Papers as First Author

1. Manabu Arikawa, Kazuhito Honda, Daisuke Akamatsu, Yoshihiko Yokoi, Keiichirou Akiba, Satoshi Nagatsuka, Akira Furusawa, and Mikio Kozuma, "Observation of electromagnetically induced transparency for a squeezed vacuum with the time domain method," Opt. Express **15**(19), 11849 (2007).

2. Manabu Arikawa, Kazuhito Honda, Daisuke Akamatsu, Satoshi Nagatsuka, Keiichirou Akiba, Akira Furusawa, and Mikio Kozuma, "Quantum memory of a squeezed vacuum for arbitrary frequency sidebands," Physical Review A **81**, 021605(R) (2010).

3. Manabu Arikawa, Takeshi Okamoto, Mitsunori Muraki, Daisaku Ogasahara, Emmanuel Le Taillandier de Gabory, Toshiharu Ito, and Kiyoshi Fukuchi, "Transmission of a 127 Gb/s PM-QPSK signal over a 3350 km SMF-only line with chromatic dispersion compensation using real-time DSP," J. Opt. Commun. Netw. **4**(11), B161 (2012).

4. Manabu Arikawa, Toshiharu Ito, Emmanuel Le Taillandier de Gabory, and Kiyoshi Fukuchi, "Crosstalk reduction with bidirectional signal assignment on square lattice structure 16-core fiber over WDM transmission for gradual upgrade of SMF-based lines," J. Light. Technol. **34**(8), 1908 (2016).

5. Manabu Arikawa and Toshiharu Ito, "Performance of mode diversity reception of a polarization-division-multiplexed signal for free-space optical communication under atmospheric turbulence," Opt. Express **26**(22), 28263 (2018).

6. Manabu Arikawa, Masaki Sato, and Kazunori Hayashi, "Wide range rate adaptation of QAM-based probabilistic constellation shaping using a fixed FEC with blind adaptive equalization," Opt. Express **28**(2), 1300 (2020).

7. Manabu Arikawa, Kohei Nakamura, Kohei Hosokawa, and Kazunori Hayashi, "Long-haul WDM/SDM transmission over coupled 4-core fiber with coupled 4-core EDFA and its mode dependent loss characteristics estimation," J. Light. Technol. **40**(6), 1664 (2022).

Other Peer-Reviewed Conference Proceedings as First Author

1. Manabu Arikawa, Takeshi Okamoto, Emmanuel Le Taillandier de Gabory, Toshiharu Ito, and Kiyoshi Fukuchi, "WDM transmission of 127 Gb/s PM-QPSK signal over 3,350 km SMF-only line with chromatic dispersion compensation using real-time DSP-LSI," OFC/NFOEC 2012, NTh1I.4.

2. Manabu Arikawa, Emmanuel Le Taillandier de Gabory, Toshiharu Ito, and Kiyoshi Fukuchi, “Improvement of signal quality after long-haul transmission over multi-core fiber with adaptive MIMO-FDE using time-domain coefficient selection,” OFC 2014, Th2A.38.
3. Manabu Arikawa, Toshiharu Ito, Emmanuel Le Taillandier de Gabory, and Kiyoshi Fukuchi, “Crosstalk reduction using bidirectional signal assignment over square lattice structure 16-core fiber for gradual upgrade of SSMF-based lines,” ECOC 2015, Th.1.2.3.
4. Manabu Arikawa, Takashi Ishikawa, Kohei Hosokawa, Seigo Takahashi, Yoshimasa Ono, and Toshiharu Ito, “Demonstration of turbulence-tolerant free-space optical communication receiver using few-mode-fiber coupling and digital combining,” IEEE Photonics Society Summer Topicals Meeting Series 2016, TuC3.4.
5. Manabu Arikawa, Takashi Ishikawa, Kohei Hosokawa, Seigo Takahashi, Yoshimasa Ono, and Toshiharu Ito, “Mitigation of fading caused by atmospheric turbulence with FMF coupling and maximum ratio combining used in 320-m free-space optical transmission of 10 Gb/s BPSK,” ECOC 2016, Th.2.P2.SC5.15.
6. Manabu Arikawa, Yoshimasa Ono, and Toshiharu Ito, “Evaluation of blind diversity combining of severely faded signals for high-speed free-space optical communication under atmospheric turbulence,” ECOC 2017, Tu.2.E.5.
7. Manabu Arikawa, Yoshimasa Ono, and Toshiharu Ito, “Mode diversity coherent receiver with few-mode fiber-coupling for high-speed free-space optical communication under atmospheric turbulence,” Free-Space Laser Communication and Atmospheric Propagation XXX (Photonics West 2018), 1052412.
8. Manabu Arikawa, “Application of optical fiber communication technologies to free-space optical communications under atmospheric turbulence,” OSA Advanced Photonics Congress 2020 (Networks), NeM4B.2.
9. Manabu Arikawa and Hidemi Noguchi, “Fast convergence by machine learning optimizer for adaptive MIMO equalizer used in SDM transmission over coupled-core 4-core fiber and 4-core EDFA,” ECOC 2020, Mo2E-6.
10. Manabu Arikawa, Kohei Hosokawa, and Kazunori Hayashi, “Long-haul WDM/SDM transmission over coupled 4-core fiber with coupled 4-core EDFA and its MDL characteristics estimation,” OFC 2021, W7D.2.
11. Manabu Arikawa, “Long-haul transmission over coupled MCF with coupled core EDFA,” OECC 2022, TuB2-4.

Bibliography

[1] K. Kikuchi, "Fundamentals of coherent optical fiber communications," J. Light. Technol. **34**(1), 157 (2016).

[2] H. Sun *et al.*, "Real-time measurements of a 40 Gb/s coherent system," Opt. Express **16**(2), 873 (2008).

[3] E. Yamazaki *et al.*, "Fast optical channel recovery in field demonstration of 100-Gbit/s Ethernet over OTN using real-time DSP," Opt. Express **19**(14), 13179 (2011).

[4] F. Derr *et al.*, "Coherent optical QPSK intradyne system: Concept and digital receiver realization," J. Light. Technol. **10**(9), 1290 (1992).

[5] S. J. Savory, "Digital filters for coherent optical receivers," Opt. Express **16**(2), 804 (2008).

[6] Y. Han and G. Li, "Coherent optical communication using polarization multiple-input-multiple-output," Opt. Express **13**(19), 7527 (2005).

[7] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," J. Light. Technol. **26**(20), 3416 (2008).

[8] I. Fatadin *et al.*, "Compensation of quadrature imbalance in an opticalQPSK coherent receiver," IEEE Photon. Technol. Lett. **20**(20), 1733 (2008).

[9] R. Rios-Müller *et al.*, "Blind receiver skew compensation and estimation for long-haul non-dispersion managed systems using adaptive equalizer," J. Light. Technol. **33**(7), 1315 (2015).

[10] H.-C. Chien *et al.*, "Single-carrier 400G PM-256QAM generation at 34 GBaud trading off bandwidth constraints and coding overheads," OFC (2017). W1J.3.

[11] F. Buchali *et al.*, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," J. Light. Technol. **34**(7), 1599 (2016).

[12] X. Chen *et al.*, "16384-QAM transmission at 10 GBd over 25-km SSMF using polarization-multiplexed probabilistic constellation shaping," ECOC (2019). PD3.3.

[13] K. Schuh *et al.*, “Single carrier 1.2 Tbit/s transmission over 300 km with PM-64QAM at 100 Gbaud,” OFC (2017). Th5B.5.

[14] M. Nakamura *et al.*, “Over 2-Tb/s net bitrate single-carrier transmission based on >130-Ghz-bandwidth InP-DHBT baseband amplifier module,” ECOC (2022). Th3C.1.

[15] H. Mardoyan *et al.*, “First 260-GBd single-carrier coherent transmission over 100 km distance based on novel arbitrary waveform generator and thin-film lithium niobate I/Q modulator,” ECOC (2022). Th3C.2.

[16] K. Ishihara *et al.*, “Frequency-domain equalisation without gurd interval for optical transmission systems,” Electron. Lett. **44**(25), 1480 (2008).

[17] L. B. Du and A. J. Lowery, “Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems,” Opt. Express **18**(16), 17075 (2010).

[18] L. Li *et al.*, “Implementation efficient nonlinear equalizer based on correlated digital backpropagation,” OFC (2011). OWW3.

[19] D. Rafique *et al.*, “Compensation of intra-channel nonlinear fibre impairments using simplified digital back-propagation algorithm,” Opt. Express **19**(10), 9453 (2011).

[20] E. Ip and J. M. Kahn, “Digital equalization of chromatic dispersion and polarization mode dispersion,” J. Light. Technol. **25**(8), 2033 (2007).

[21] C. R. S. Fludger *et al.*, “Coherent equalization and POLMUX-RZ-DQPSK for robust 100-GE transmission,” J. Light. Technol. **26**(1), 64 (2008).

[22] D. N. Godard, “Self-recovering equalization and carrier tracking in two-dimensional data communication systems,” IEEE Trans. on Commun. **COM-28**(11), 1867 (1980).

[23] D.-S. Ly-Gagnon *et al.*, “Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation,” J. Light. Technol. **24**(1), 12 (2006).

[24] A. Leven *et al.*, “Frequency estimation in intradyne reception,” IEEE Photon. Technol. Lett. **19**(6), 366 (2007).

[25] Y. LeCun *et al.*, “Deep learning,” Nature **521**, 436 (2015).

[26] T. Tanimura *et al.*, “Convolutional neural network-based optical performance monitoring for optical transport networks,” J. Opt. Commun. Netw. **11**(1), A52 (2019).

[27] D. Zibar *et al.*, “Machine learning-based Raman amplifier design,” OFC (2019). M1J.1.

[28] B. Karanov *et al.*, “End-to-end deep learning of optical fiber communications,” J. Light. Technol. **36**(20), 4843 (2018).

[29] S. Gaiarin *et al.*, “End-to-end optimization of coherent optical communications over the split-step Fourier method guided by the nonlinear Fourier transform theory,” J. Light. Technol. **39**(2), 418 (2021).

[30] C. Häger and H. D. Pfister, “Physics-based deep learning for fiber optic communication systems,” IEEE J. Sel. Areas Commun. **39**(1), 280 (2021).

[31] B. I. Bitachon *et al.*, “Deep learning based digital backpropagation demonstrating SNR gain at low complexity in 1200 km transmission link,” Opt. Express **28**(20), 29318 (2020).

[32] T. T. Nguyen *et al.*, “Coupled transceiver-fiber nonlinearity compensation based on machine learning for probabilistic shaping system,” J. Light. Technol. **39**(2), 388 (2020).

[33] C. Bluemm *et al.*, “Equalizing nonlinearities with memory effects: Volterra series vs. deep neural networks,” ECOC (2019). W.3.B.3.

[34] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” Math. Control. Signals Syst. **2**, 303 (1989).

[35] S. Amari *et al.*, “Backpropagation and stochastic gradient descent method,” Neurocomputing **5**, 185 (1993).

[36] A. G. Baydin *et al.*, “Automatic differentiation in machine learning: a survey,” J. Mach. Learn. Res. **18**, 1 (2018).

[37] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” Int. Conf. on Mach. Learn. (2010). 399.

[38] S. Takabe *et al.*, “Trainable projected gradient detector for massive overloaded MIMO channels: Data-driven tuning approach,” IEEE Access **7**, 93326 (2019).

[39] M.-W. Un *et al.*, “Deep MIMO detection using ADMM unfolding,” Data Sci. Work. (2019). 333.

[40] S. Haykin, *Adaptive filter theory fourth ed.* (Prentice Hall, 2002).

[41] J. G. Proakis and M. Salehi, *Digital communications fifth edition* (McGraw-Hill, 2008).

[42] E. P. da Silva *et al.*, “Widely linear equalization for IQ imbalance and skew compensation in optical coherent receivers,” J. Light. Technol. **34**(15), 3577 (2016).

[43] C. E. Shannon, “Communication in the presence of noise,” Proc. IRE **37**(1), 10 (1949).

[44] P. Poggiolini *et al.*, “The GN-model of fiber non-linear propagation and its applications,” J. Light. Technol. **32**(4), 694 (2014).

[45] A. Yariv and P. Yei, *Photonics: optical electronics in modern communications sixth edition* (Oxford university, 2007).

[46] G. P. Agrawal, *Nonlinear fiber optics third ed.* (Academic press, 2001).

[47] C. Antonelli *et al.*, “Modeling of nonlinear propagation in space-division multiplexed fiber-optic transmission,” J. Light. Technol. **34**(1), 36 (2016).

[48] S. Zhang *et al.*, “Field and lab experimental demonstration of nonlinear impairment compensation using neural networks,” Nat. Comm. **10**, 3033 (2019).

[49] R.-J. Essiambre *et al.*, “Capacity limits of optical fiber networks,” J. Light. Technol. **28**(4), 662 (2010).

[50] M. S. Faruk *et al.*, “Multi-impairment monitoring from adaptive finite-impulse-response filters in a digital coherent receiver,” Opt. Express **18**(26), 26929 (2010).

[51] L. E. Nelson *et al.*, “Field measurements of polarization transients on a long-haul terrestrial link,” IEEE Photonic Soc. Annu. Meet. (2011). ThT5.

[52] P. M. Krummrich *et al.*, “Demanding response time requirements on coherent receivers due to fast polarization rotations caused by lightning events,” Opt. Express **24**(11), 12242 (2016).

[53] Y. Yoshida *et al.*, “Single-pixel optical modulation analyzer: a low-complexity frequency-dependent IQ imbalance monitor based on direct detection with phase retrieval,” Opt. Express **29**(22), 31115 (2021).

[54] C. R. S. Fludger and T. Kupfer, “Transmitter impairment mitigation and monitoring for high baud-rate, high order modulation systems,” ECOC (2016). Tu.2.A.2.

[55] T. Pfau *et al.*, “Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for m-QAM constellations,” J. Light. Technol. **27**(8), 989 (2009).

[56] Q. Zhang *et al.*, “Algorithms for blind separation and estimation of transmitter and receiver IQ imbalances,” J. Light. Technol. **37**(10), 2201 (2019).

[57] Y. Fan *et al.*, “Transceiver IQ imperfection monitor by digital signal processing in coherent receiver,” OECC/PSC (2019). ThC2-1.

[58] C. Ju *et al.*, “In-service blind transceiver IQ imbalance and skew monitoring in long-haul non-dispersion managed coherent optical systems,” IEEE Access **7**, 150051 (2019).

[59] T. Kobayashi *et al.*, “35-Tb/s C-band transmission over 800 km employing 1-Tb/s PS-64QAM signals enhanced by complex 8 × 2 MIMO equalizer,” OFC (2019). Th4B.2.

[60] M. Sato *et al.*, “Transceiver impairment mitigation by 8×2 widely linear MIMO equalizer with independent complex filtering on IQ signals,” submitted to IEEE Photon. J. .

[61] R. Kudo *et al.*, “Two-stage overlap frequency domain equalization for long-haul optical systems,” OFC/NFOEC (2009). OMT3.

[62] L. E. Nelson *et al.*, "A robust real-time 100G transceiver with soft-decision forward error correction," J. Opt. Commun. Netw. **4**(11), B131 (2012).

[63] I. Goodfellow *et al.*, *Deep Learning* (MIT Press, 2016).

[64] Y. Mori *et al.*, "Novel configuration of finite-impulse-response filters tolerant to carrier-phase fluctuations in digital coherent optical receivers for higher-order quadrature amplitude modulation signals," Opt. Express **20**(24), 26236 (2012).

[65] W. Shieh and K.-P. Ho, "Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing," Opt. Express **16**(20), 15718 (2008).

[66] T. Xu *et al.*, "Equalization enhanced phase noise in Nyquist-spaced superchannel transmission systems using multi-channel digital back-propagation," Sci. Reports **5**, 13990 (2015).

[67] J. Yang *et al.*, "The multimodulus blind equalization and its generalized algorithms," IEEE J. on Sel. Areas Commun. **20**(5), 997 (2002).

[68] C. R. Johnson *et al.*, "Blind equalization using the constant modulus criterion: A review," Proc. The IEEE **86**(10), 1927 (1998).

[69] S. Randel *et al.*, "Complexity analysis of adaptive frequency-domain equalization for MIMO-SDM transmission," ECOC (2013). Th.2.C.4.

[70] M. Mazur *et al.*, "Overhead-optimization of pilot-based digital signal processing for flexible high spectral efficiency transmission," Opt. Express **27**(17), 24654 (2019).

[71] Y. Wakayama *et al.*, "Increasing achievable information rates with pilot-based DSP in standard intradyne detection," ECOC (2019). W.1.B.6.

[72] M. Selmi *et al.*, "Accurate digital frequency offset estimator for coherent PolMux QAM transmission systems," ECOC (2009). P3.08.

[73] I. Fatadin *et al.*, "Compensation of frequency offset for differentially encoded 16- and 64-QAM in the presence of laser phase noise," IEEE Photon. Technol. Lett. **22**(3), 176 (2010).

[74] J. Cho *et al.*, "Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM," ECOC (2017). M.2.D.2.

[75] D. C. Kilper, "Optical performance monitoring," J. Light. Technol. **22**(1), 294 (2004).

[76] A. Matsushita *et al.*, "High-spectral efficiency 600-Gb/s/carrier transmission using PDM-256QAM format," J. Light. Technol. **37**(2), 470 (2019).

[77] M. N. Petersen *et al.*, "Online chromatic dispersion monitoring and compensation using a single inband subcarrier tone," IEEE Photon. Technol. Lett. **14**(4), 570 (2002).

[78] K. J. Park *et al.*, "Performance comparison of chromatic dispersion-monitoring techniques using pilot tones," IEEE Photon. Technol. Lett. **15**(6), 873 (2003).

[79] B. Fu and R. Hui, "Fiber chromatic dispersion and polarization-mode dispersion monitoring using coherent detection," IEEE Photon. Technol. Lett. **17**(7), 1561 (2005).

[80] J. Yang *et al.*, "CD-insensitive PMD monitoring based on RF power measurement," Opt. Express **19**(2), 1354 (2011).

[81] C. Wang *et al.*, "Joint OSNR and CD monitoring in digital coherent receiver using long short-term memory neural network," Opt. Express **27**(5), 6936 (2019).

[82] F. N. Hauske *et al.*, "Optical performance monitoring from FIR filter coefficients in coherent receivers," OFC/NFOEC (2008). OThW2.

[83] J. Liang *et al.*, "Transceiver imbalances compensation and monitoring by receiver DSP," J. Light. Technol. **39**(17), 5397 (2021).

[84] A. Kawai *et al.*, "Digital four-dimensional transceiver IQ characterization," J. Light. Technol. **41**(5), 1389 (2023).

[85] K.-P. Ho, *Phase-modulated optical communication systems* (Springer, 2005).

[86] F. M. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers," IEEE Trans. on Commun. **COM-34**(5), 423 (1986).

[87] F. Buchali *et al.*, "Beyond 100 Gbaud transmission supported by a 120 GSa/s CMOS digital to analog converter," ECOC (2019). Tu.2.D.3.

[88] M. Nakamura *et al.*, "1.3-Tbps/carrier net-rate signal transmission with 168-Gbaud PDM PS-64QAM using analogue-multiplexer-integrated optical frontend module," ECOC (2019). Tu.2.D.5.

[89] M. R. Chitgarha *et al.*, "2×800Gbps/wave coherent optical module using a monolithic InP tranceiver PIC," ECOC (2020). We2C-1.

[90] A. Rezania *et al.*, "Compensation schemes for transmitter- and receiver-based pattern-dependent distortion," IEEE Photon Technol. Lett. **28**(22), 2641 (2016).

[91] G. Khanna *et al.*, "A robust adaptive pre-distortion method for optical communication transmitters," IEEE Photon Technol. Lett. **28**(7), 752 (2016).

[92] P. W. Berenguer *et al.*, "Nonlinear digital pre-distortion of transmitter components," J. Light. Technol. **34**(8), 1739 (2016).

[93] H. Faig *et al.*, "Dimensions-reduced Volterra digital pre-distortion based on orthogonal basis for band-limited nonlinear opto-electronic components," IEEE Photon. J. **11**(1), 7201013 (2019).

[94] M. Abu-Romoh *et al.*, “Neural-network-based pre-distortion method to compensate for low resolution DAC nonlinearity,” ECOC (2019). Th.1.B.4.

[95] G. Paryanti *et al.*, “A direct learning approach for neural network based pre-distortion for coherent nonlinear optical transmitter,” J. Light. Technol. **38**(15), 3883 (2020).

[96] V. Bajaj *et al.*, “Single-channel 1.61 Tb/s optical coherent transmission enabled by neural network-based digital pre-distortion,” ECOC (2020). Tu1D-5.

[97] E. Ip *et al.*, “Coherent detection in optical fiber systems,” Opt. Express **16**(2), 753 (2008).

[98] F. Pittalà and J. A. Nossek, “800ZR+ DWDM demonstration over 600km G.654D fiber enabled by adaptive nonlinear tripleX equalization,” OFC (2020). M4K.5.

[99] J. Tsimbinos and K. V. Lever, “The computational complexity of nonlinear compensators based on the Volterra inverse,” 8th Work. on Stat. Signal Array Process. (1996). 387.

[100] V. J. Mathews, “Adaptive polynomial filters,” IEEE Signal Process. Mag. **8**(3), 10 (1991).

[101] S. Amiralizadeh *et al.*, “Modeling and compensation of transmitter nonlinearity in coherent optical OFDM,” Opt. Express **23**(20), 26192 (2015).

[102] O. Sidelnikov *et al.*, “Advanced convolutional neural network for nonlinearity mitigation in long-haul WDM transmission systems,” J. Light. Technol. **39**(8), 2397 (2021).

[103] R. M. Bütler *et al.*, “Model-based machine learning for joint digital backpropagation and PMD compensation,” J. Light. Technol. **39**(4), 949 (2021).

[104] P. Chevalier *et al.*, “Complex transversal volterra filters optimal for detection and estimation,” ICASSP 1991), E8.2.

[105] T. A. Eriksson *et al.*, “Applying neural networks in optical communication systems: Possible pitfalls,” IEEE Photon. Technol. Lett. **29**(23), 2091 (2017).

[106] P. J. Freire *et al.*, “Neural networks-based equalizers for coherent optical transmission: Caveats and pitfalls,” J. Sel. Top. Quantum Electron. **28**(4), 7600223 (2022).

[107] X. Chen *et al.*, “Experimental quantification of implementation penalties from laser phase noise for ultra-high-order QAM signals,” ECOC (2020). Tu2D-5.

[108] Q. Zhuge *et al.*, “Pilot-aided carrier phase recovery for m-QAM using superscalar parallelization based PLL,” Opt. Express **20**(17), 19599 (2012).

[109] J. Cho, “Balancing probabilistic shaping and forward error correction for optimal system performance,” OFC (2018). M3C.2.

[110] J.-X. Cai *et al.*, “70.46 Tb/s over 7,600 km and 71.65 Tb/s over 6,970 km transmission in C+L band using coded modulation with hybrid constellation shaping and nonlinearity compensation,” J. Light. Technol. **36**(1), 114 (2018).

[111] M. F. C. Stephens *et al.*, “Trans-atlantic real-time field trial using super-Gaussian constellation-shaping to enable 30Tb/s+ capacity,” OFC (2021). F4G.2.

[112] G. Böcherer *et al.*, “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” IEEE Trans. on Commun. **63**(12), 4651 (2015).

[113] J. Cho and P. J. Winzer, “Probabilistic constellation shaping for optical fiber communications,” J. Light. Technol. **37**(6), 1590 (2019).

[114] P. Schulte and G. Böcherer, “Constant composition distribution matching,” IEEE Trans. on Info. Theory **62**(1), 430 (2016).

[115] S. Randel *et al.*, “6×56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6×6 MIMO equalization,” Opt. Express **19**(17), 16697 (2011).

[116] P. J. Winzer, “Challenges and evolution of optical transport networks,” ECOC (2010). We.8.D.1.

[117] G. Bosco *et al.*, “On the performance of Nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers,” J. Light. Technol. **29**(1), 53 (2011).

[118] D. J. Richardson *et al.*, “Space-division multiplexing in optical fibres,” Nat. Photonics **7**, 354 (2013).

[119] P. J. Winzer *et al.*, “Fiber-optic transmission and networking: the previous 20 and the next 20 years,” Opt. Express **28**(18), 24190 (2018).

[120] V. A. J. M. Sleiffer *et al.*, “73.7 Tb/s (96 × 3 × 256-gb/s) mode-division-multiplexed DP-16QAM transmission with inline MM-EDFA,” Opt. Express **20**(26), B428 (2012).

[121] D. Soma *et al.*, “50.47-Tbit/s standard cladding coupled 4-core fiber transmission over 9,150 km,” J. Light. Technol. **39**(22), 7099 (2021).

[122] M. Mazur *et al.*, “Real-time MIMO transmission over field-deployed coupled-core multi-core fibers,” OFC (2022). Th4B.8.

[123] T. Hayashi *et al.*, “Record-low spatial mode dispersion and ultra-low loss coupled multi-core fiber for ultra-long-haul transmission,” J. Light. Technol. **35**(3), 450 (2017).

[124] R. Ryf *et al.*, “Long-haul transmission over multi-core fibers with coupled cores,” ECOC (2017). M.2.E.1.

[125] R. Ryf *et al.*, “Coupled-core transmission over 7-core fiber,” OFC (2019). Th4B.3.

[126] S. Beppu *et al.*, “Real-time strongly-coupled 4-core fiber transmission,” OFC (2020). Th3H.2.

[127] S. Beppu *et al.*, “Long-haul coupled 4-core fiber transmission over 7,200 km with real-time MIMO DSP,” J. Light. Technol. **40**(6), 1640 (2022).

[128] M. Mazur *et al.*, “Comparison of transfer matrix stability between a 110 km 7-core coupled-core multi-core fiber and single-mode fiber,” OFC (2022). M1E.2.

[129] C. Koebele *et al.*, “40km transmission of five mode division multiplexed data streams at 100Gb/s with low MIMO-DSP complexity,” ECOC 2011). Th.13.C.3.

[130] S. Beppu *et al.*, “402.7-Tb/s MDM-WDM transmission over weakly-coupled 10-mode fiber using rate-adaptive PS-16QAM signals,” J. Light. Technol. **38**(10), 2835 (2020).

[131] K. Shibahara *et al.*, “Dense SDM (12-core × 3-mode) transmission over 527 km with 33.2-ns mode-dispersion employing low-complexity parallel MIMO frequency-domain equalization,” J. Light. Technol. **34**(1), 196 (2016).

[132] J. J. Shynk, “Frequency-domain and multirate adaptive filtering,” IEEE Signal Process. Mag. **9**(1), 14 (1992).

[133] M. S. Faruk and K. Kikuchi, “Adaptive frequency-domain equalization in digital optical receivers,” Opt. Express **19**(13), 12789 (2011).

[134] N. Bai and G. Li, “Adaptive frequency-domain equalization for mode-division multiplexed transmission,” IEEE Photon. Technol. Lett. **24**(21), 1918 (2012).

[135] K. Kikuchi, “Clock recovering characteristics of adaptive finite-impulse-response filters in digital coherent optical receivers,” Opt. Express **19**(6), 5611 (2011).

[136] S. F. A. Shah *et al.*, “Low-complexity design of noninteger fractionally spaced adaptive equalizers for coherent optical receivers,” IEEE Signal Process. Lett. **23**(9), 1289 (2016).

[137] C. Li *et al.*, “Advanced DSP for single-carrier 400-Gb/s PDM-16QAM,” OFC (2016). W4A.4.

[138] F. Hamaoka *et al.*, “Experimental demonstration of simplified adaptive equalizer for fractionally sampled 120-gbaud signal,” ECOC (2017). P1.SC3.34.

[139] F. Pittalà and J. A. Nossek, “Efficient frequency-domain fractionally-spaced equalizer for flexible digital coherent optical receivers,” ECOC (2015). P.3.14.

[140] M. Paskov *et al.*, "A fully-blind fractionally-oversampled frequency domain adaptive equalizer," OFC (2016). Th2A.33.

[141] P. P. Vaidyanathan, *Multirate systems and filter banks* (Prentice Hall, 1993).

[142] M. Arikawa *et al.*, "Transmission of a 127 Gb/s PM-QPSK signal over a 3350 km SMF-only line with chromatic dispersion compensation using real-time DSP," J. Opt. Commun. Netw. **4**(11), B161 (2012).

[143] P. Duhamel and M. Vetterli, "Fast Fourier transforms: A tutorial review and a state of the art," Signal Process. **19**, 259 (1990).

[144] C. S. Burrus and P. W. Eschenbacher, "An in-place, in-order prime factor FFT algorithm," Trans. on Acoust. Speech, Signal Process. **ASSP-29**(4), 806 (1981).

[145] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," J. Sel. Top. Quantum Electron. **16**(5), 1180 (2010).

[146] L. Liu *et al.*, "Initial tap setup of constant modulus algorithm for polarization de-multiplexing in optical coherent receivers," OFC/NFOEC (2009). OMT2.

[147] Z. Y. Choi and Y. H. Lee, "Frame synchronization in the presence of frequency offset," IEEE Trans. on Commun. **50**(7), 1062 (2002).

[148] T. Tanimura *et al.*, "Fiber-longitudinal anomaly position identification over multi-span transmission link out out receiver-end signals," J. Light. Technol. **38**(9), 2726 (2020).

[149] T. Sasai *et al.*, "Digital longitudinal monitoring of optical fiber communication link," J. Light. Technol. **40**(8), 2390 (2022).