MODAL EFFECT CHARACTERIZATION AND COMPENSATION FOR HIGH-CAPACITY, SPACE-DIVISION-MULTIPLEXED OPTICAL NETWORKS

by

Mahmoudreza Dadras Jedi Goldehi

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Electrical Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2025

DEDICATION

To my dear wife, your love, support and the many sacrifices you have made have been my greatest strength. Even in the toughest moments, your belief in me kept me going. I couldn't have done this without you, and this achievement is just as much yours as it is mine. To my mom, whose love and support have been my source of strength. To my siblings, friends, and mentors, for their endless motivation and support. To the loving memory of my father, who never witnessed this journey, but whose presence was felt every step of the way.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Prof. Ioannis Roudas, for his invaluable guidance, unwavering support, and insightful feedback throughout the course of this research. His expertise and encouragement have played a crucial role in shaping this dissertation and expanding my knowledge in the field of optical communications.

I extend my sincere appreciation to my committee members, Prof. Jessica Jiang, Prof. Jaroslaw Kwapisz, Prof. David Dickensheets, and Prof. Kevin Repasky, for their guidance and valuable suggestions that have greatly enriched the quality of this work.

I am also grateful to my group members at MSU Optical Communications group for their collaboration and shared knowledge, which have helped me in this journey.

Finally, I would like to acknowledge the financial support provided by National Science Foundation (NSF), without which this research would not have been possible.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF ALGORITHMS

NOMENCLATURE

SDM	Spatial division multiplexing
MCF	Multicore fiber
MMF	Multimode fiber
SMF	Single mode fiber
DSP	Digital signal processing
MD-SDM	Mode-dependent signal delay method
MD-APM	Mode-dependent average power method
PD-SDM	Polarization-dependent signal delay method
SWI	Swept-wavelength interferometry
SNR	Signal-to-noise ratio
MD	Modal dispersion
CD	Chromatic dispersion
PMD	Polarization mode dispersion
PM	Principal modes
MDL	Mode dependent loss
PDL	polarization dependent loss
DMGD	Differential mode group delay
ISI	Intersymbol interference
MIMO	Multiple-input multiple-output
ROADM	Reconfigurable optical add-drop multiplexer
SM	Spatial multiplexer
SD	Spatial demultiplexer
IL	Insertion loss
LP	Linearly polarized
DGD	Differential group delay
SLM	Spatial light modulator
SVM	Stokes vector modulation
SVR	Stokes vector receiver
MVM	Mode vector modulation
OOK	On-off keying
DMT	Discrete multitone
EML	Electro-absorption modulation
IMDD	Intensity modulation direct detection
PAM	Pulse amplitude modulation
KK	Kramers-Kronig
SPPM	Spatial pulse position modulation
HD-SVDD	High-dimensional Stokes vector direct detection
SOP	State of the polarization
RSOP	Rotation of the state of the polarization
CD	Chromatic dispersion

NOMENCLATURE – CONTINUED

WDM	Wavelength-division multiplexing
DWDM	Dense wavelength-division multiplexing
BER	Bit error rate
SER	Symbol error rate
LMS	Least mean square
ICA	Independent component analysis
CMA	Constant modulus algorithm
EKF	Extended Kalman filter
OEKF	Optimized extended Kalman filter
BPS	Blind phase search
SWA-EKF	Sliding window-aided extended Kalman filter
EEPN	Equalization-enhanced phase noise
PS	Probabilistically shaped

ABSTRACT

Due to the exponential growth of global data traffic, a capacity shortage in the optical fiber network is expected in the near future. Space-division multiplexing (SDM), combined with multicore fibers (MCFs) and multimode fibers (MMFs), offers a promising solution to address this challenge. For this technology to become commercially viable, issues such as MCF and MMF characterization and the impact of transmission impairments on SDM optical links must be addressed. In addition, since multiple modes or cores can be jointly utilized to transmit information, energy-efficient multi-dimensional modulation formats and the corresponding transceivers, along with their digital signal processing (DSP) algorithms, should also be considered.

First, we evaluate two methods, namely the mode-dependent signal delay method (MD-SDM) and the mode-dependent average power method (MD-APM), for the characterization of MCFs and MMFs. In these methods, we select launch states that optimize the signal-to-noise ratio (SNR) for measuring the modal dispersion (MD) and mode-dependent loss (MDL) vectors. The MD of MMFs and MCFs is estimated using time-of-flight measurements of optical pulses and multiple sets of group delay measurements. We then investigate the accuracy of MD-SDM and MD-APM in characterizing MMFs by simulating various sources of error. Additionally, we estimate the MD and MDL of MMFs and MCFs under different coupling regimes. Thermal noise at the direct-detection receiver and modal crosstalk at the fiber input are considered as two limiting factors in the estimation of the MD and MDL vectors.

During the second phase of this dissertation, we study various multi-dimensional modulation formats for MMFs and MCFs. Our focus is on mode vector modulation (MVM), which was recently proposed by our group. We model and simulate transmission impairments such as rotation of the state of polarization (RSOP), MD, chromatic dispersion (CD), and modal crosstalk in SDM wavelength-division multiplexing (WDM) optical networks. Next, we investigate different DSP equalization algorithms, including least mean square (LMS), independent component analysis (ICA), Kalman filter, and extended Kalman filter (EKF). Finally, we propose an optimal adaptive equalizer for the MVM direct-detection transceiver.

INTRODUCTION

The proliferation of data traffic over the internet may lead to a capacity shortage in fiber-optic networks, driven by emerging technologies that require increased processing, storage, and transmission of information — such as 5G applications, cloud computing, the Internet of Things (IoT), and artificial intelligence (AI). In addition, as shown in Fig. 1, and according to the Nokia Global Network Traffic Report [1], overall network traffic is expected to grow by 5 to 9 times by 2033, rising from 396 EB per month in 2023 to 3,280 EB per month in 2033.

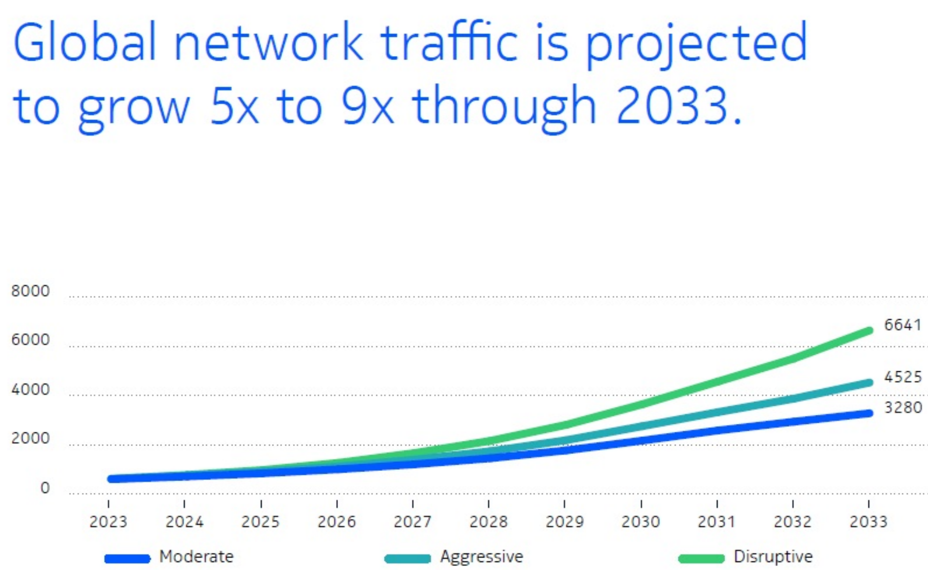

Figure 1: Global IP traffic growth [1].

Optical communication systems form the backbone of the global internet and play a crucial role in meeting human information transmission needs. Both long-haul data transmission between countries and continents, as well as short-range optical links in data centers — the main focus of this project — rely on optical communications. Therefore, it is essential to enhance the capacity of optical communication systems to keep pace with growing global data traffic demands.

An approach to improve the optical networking infrastructure: space division multiplexing

Space-division multiplexing (SDM), in conjunction with multimode fibers (MMFs) and multicore fibers (MCFs), can significantly increase the capacity of fiber-optic systems [3]. Different spatial paths, provided by the modes in MMFs and the cores in MCFs, can be utilized to transmit independent data streams [4]. This approach offers a substantial capacity upgrade that can help avert the predicted capacity crunch, at least in the foreseeable future. In addition, the use of SDM technology leads to improvements in spectral efficiency, as well as reductions in both cost and energy consumption [5]. Several research groups have demonstrated that by combining wavelength-division multiplexing (WDM) with SDM over MMFs and MCFs, it is possible to achieve aggregate data rates exceeding 10 Pb/s [6], [7]. Fig. 2 illustrates the evolution of WDM and SDM data rates in optical fiber communication systems. Recently, it has been shown that WDM systems can achieve 178.08 Tb/s over 40 km [8] and 206.1 Tb/s over 54 km [2] using single-mode fiber. Therefore, SDM technology can be employed to transmit even higher data rates by utilizing parallel spatial channels.

Characterization of SDM Systems, transmission impairments, and components

Transmission impairments in MMFs and MCFs — such as modal dispersion (MD), mode-dependent loss (MDL), and intermodal/intercore coupling — must be carefully considered in the design of a reliable and efficient multiple-input multiple-output (MIMO) system. MD is the dominant limiting factor in uncompensated MMF links, as it induces severe intersymbol interference (ISI) at the optical receiver [9], [10], and thus must be

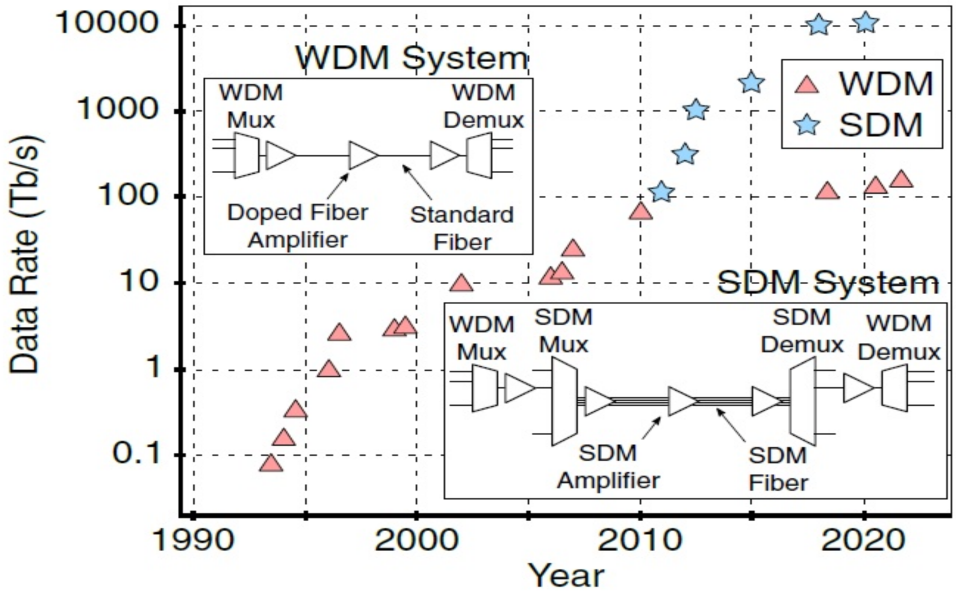

Figure 2: Evolution of data rates for WDM and SDM systems [2].

properly characterized. The formalism of polarization mode dispersion (PMD) in single-mode fibers (SMFs) can be generalized to describe MD in MMFs. Specifically, MD can be characterized by the principal modes (PMs) and their corresponding differential mode group delays (DMGDs), which are represented as a vector in a generalized Stokes space — the so-called MD vector [11].

Several methods have been proposed in the optical communications literature to measure the MD of MMFs. For instance, Ahn et al. [12] introduced a technique based on an intermodal interferometer combined with optical frequency-domain reflectometry. In [13], the modal content of the fiber was characterized by spectrally and spatially resolving the fiber output at different wavelengths. Imaging-based methods were also proposed in [13], relying on the same principle of spectral and spatial imaging. Demas et al. [14] employed a similar setup to that in [13], with the main difference being that the reference and signal paths were separated. Carpenter et al. [15] utilized swept-wavelength interferometry (SWI) to measure the fiber transfer matrix. Unlike the previously mentioned techniques, SWI requires a coherent optical receiver but offers a more comprehensive characterization of MMFs. Specifically, SWI can simultaneously provide information on MD, PMs, DMGDs, and MDL of the fiber under test, as it relies on amplitude and phase measurements of the input and output modes to construct the fiber transfer matrix.

A recently proposed method that can fully characterize the MD of a MMF is the mode-dependent signal delay method (MD-SDM) [16]. This method is a generalization of the polarization-dependent signal delay method (PD-SDM), which is used to measure the PMD vector in SMFs [16]. In MD-SDM, the time-of-flight of optical pulses is employed to characterize the MD of the fiber, a process that can be performed using a direct-detection receiver. It has been analytically shown [17] that the components of the input MD vector can be determined through such group delay measurements. Once the input MD vector is known, the input group-delay operator can be constructed, allowing for the estimation of both the input MD and the DMGDs [17]. Similarly, for MDL characterization, one can determine the MDL vector by launching optical pulses and measuring their average output powers. We refer to this characterization technique as the MD-APM. To implement an SDM transmission system, a combination of SMF and SDM-specific components must be used. These components include spatial multiplexers and demultiplexers, photonic lanterns, SDM reconfigurable optical add-drop multiplexers (ROADMs), and SDM amplifiers, all of which need to be generalized for SDM operation [18], [19], [20], [21].

Spatial Multiplexers

One of the most important components in an SDM system is the spatial multiplexer (SM) and spatial demultiplexer (SD). These devices are designed to select and manipulate various propagation modes. An SM typically contains a mode converter that can transform input signals propagating in a given spatial mode, such as the fundamental linearly polarized mode (LP_{01}), into other modes. Coupling between individual modes and mode groups is a critical factor to consider when using SMs and SDs.

The transfer matrix of an SM is typically diagonal and contains key information such as insertion loss (IL) and MDL. Mode-selective SMs and mode-group-selective SMs serve a variety of functions in SDM systems. Notably, mode-group-selective SMs can excite specific mode combinations in the fiber and compensate for differential group delays (DGDs) among mode groups [22].

Phase mask SMs, such as spatial light modulators (SLMs), are popular choices for SDM communication systems and are also widely used in component characterization [23], [24].

Photonic Lanterns

To merge signals from a SMF bundle and launch them into different modes of a MMF, a mode converter such as a photonic lantern is required. This low-insertion-loss and low-MDL SDM component can generate various mode profiles, thereby enhancing the ability to excite higher-order spatial modes. Modal selectivity and the generation of a larger number of modes are achieved through specialized microstructured designs and particular geometrical arrangements [25].

SDM ROADMs

Wavelength management (or optical channel/carrier management) is critical for optimizing the optical spectrum in a core optical network using dense wave division multiplexing (DWDM). Key aspects of wavelength management include optimizing the wavelength plan of the optical spectrum through flexible optical carrier center frequency and bandwidth provisioning, improving configuration efficiency and speed via remote management and provisioning, and enhancing optical performance by managing individual optical carrier power levels at various points in the network. Therefore, a network element with ROADM functionality (ROADM NE) provides a powerful solution for such optical networks.

An SDM ROADM can manage signals propagating over multiple spatial paths while routing and switching wavelengths. In fact, an SDM ROADM must reduce service costs while maintaining equal or superior performance compared to SMF systems. Various SDM ROADM architectures have been proposed, such as 2-port ROADMs for 7-core fibers and FMF ROADMs with and without spatial diversity [26], [27], [28], [29], [30].

To deploy an SDM system, commercially available WDM systems must be leveraged, as current optical networks predominantly rely on this technology. Since the ROADM is the heart of an optical network, it must support a flex-grid spectrum in an SDM system while enabling switching between cores or modes — a very challenging requirement. Consequently, the switching capability of SDM ROADMs is a critical factor. For example, a 12-port WDM ROADM with 96 channels would require 1,152 switching states. Therefore, in SDM/WDM ROADMs, spatial channels must also be handled, which significantly increases switching complexity. To address this challenge, joint switching can be employed to reduce switching complexity [28], [31]. This joint switching should be implemented using one of the steering elements within a WSS.

Modulation formats in inter-data-center networks: past, present, and future

Optical interconnects are widely deployed in data centers worldwide as the demand for higher link capacity continues to grow. In data centers, intensity modulation direct detection (IMDD) has been adopted to reduce costs and power consumption. The spectral efficiency of these optical interconnects can be improved through modulation formats such as M-ary pulse amplitude modulation (PAM), an extension of binary on-off keying (OOK) with multiple power levels [32], [33], [34], [35].

Recently, Zhang et al. [36] demonstrated the transmission of an 800 Gbit/s signal using four channels (4×200 Gbit/s), which was later confirmed by Okuda et al., who achieved 800 Gbit/s using PAM-4 transceivers operating at 53.125 Gbaud [37]. A review by Pang et al. [38] provides an overview of recent advancements in IMDD transmission, along with key technologies aimed at achieving 200 Gbit/s per lane using PAM-4 and discrete multitone (DMT) formats. Additionally, as reported in [39], more than 200 Gbit/s DMT signals were successfully transmitted over 10 km using an electro-absorption modulation laser (EML)-based direct-detection system.

Coherent optical receivers have the potential to enhance the performance and increase the capacity of optical connections using SMFs as an alternative to direct-detection systems. Perin et al. [40] compared various receiver architectures, including DSP-free coherent and Kramers-Kronig (KK) receivers. Their study found that conventional DSP-based coherent receivers, originally designed for long-haul transmission, are not suitable for short-reach applications. Instead, simplified and lower-cost coherent receiver designs may be more appropriate for short links. Therefore, the research concluded that standard DSP-based coherent receivers developed for long-distance transmission are impractical for short-reach systems, and future efforts should focus on making coherent links more attractive by adopting simpler receiver architectures.

Researchers from MICRAM Microelectronic and Nokia Bell Labs demonstrated transmission rates of 1.55 Tb/s net bit rate in back-to-back configuration, 1.52 Tb/s after 80 km, and 1.46 Tb/s after 240 km of SMF [41]. Coherent optical communication solutions have become widely adopted in long-reach fiber transmission systems, enabled by technologies such as optical amplifiers and ROADMs.

To fulfill the ever-increasing demand of short-reach optical networks, multidimensional modulation formats have been developed, leveraging the superior spectral efficiency and sensitivity of coherent optical systems compared to direct detection [42], [43], [44], [45], [46]. These additional degrees of freedom are enabled by exploiting different signal levels, polarization states, and spatial modes. Various DSP techniques have also been employed to mitigate impairments associated with coherent detection [47], [48], [43], [49]. Moreover, the success of coherent detection systems has stimulated intensive research into new modulation formats and constellation shaping optimization, all aiming to enhance the utility of these formats [50], [51]. Constellation shaping has been a natural outcome of source optimization, with the goal of maximizing the mutual information between the input and output of the channel [52], [53], [54]. One of the techniques that has recently gained significant interest is Stokes vector modulation (SVM) direct detection (DD), which has the potential to bridge the gap between coherent detection systems and IM/DD systems. Notably, SVM exploits both polarization states to increase capacity per wavelength, making it highly suitable for short-range interconnects [55].

SVM relies on phase shifts between polarization states, and three dimensions of the Stokes space can be exploited to create suitable constellations. Although, in principle, SVM may reduce receiver complexity compared to coherent detection systems, it can compromise sensitivity due to the limitations of direct detection [56], [57], [58], [59], [60], [61], [62].

Another technique that can also be used in creating multi-dimensional modulation formats, especially with coherent detection in optical networks, is spatial pulse position modulation (SPPM). In the place of time slots used in regular PPM, multiple cores or modes can be used [63]-[64], [65]. Plant et al. [66] reviewed multidimensional modulation formats and explained the advantages of using these modulations with direct-detection receivers.

Furthermore, as SDM gains traction in optical networks, there is a growing need for modulation formats capable of addressing multidimensional data transmission requirements. In SVM, variations in the state of polarization (SOP) are induced by time-varying birefringence in the fiber. In coherent detection, the SOP can be characterized using a 2×2 Jones matrix [67].

For the purpose of enhancing capacity and power efficiency, multidimensional SVM becomes essential. By utilizing additional dimensions, SVM demonstrates higher efficiency compared to other formats such as M-PAM. Experimental evidence supporting the shift to higher-dimensional SVM is presented in [68], where high-dimensional Stokes vector direct detection (HD-SVDD) achieved the transmission of a 60 Gb/s signal over a 200-meter few-mode fiber using four spatial and polarization states of the LP_{11} mode group. Moreover, a new framework for generalized SVM, known as mode vector modulation (MVM), has recently been proposed to enable optimal transceiver architectures [69], [70], [71].

Digital signal processing receiver for multi-dimensional modulation format

In this new high-dimensional format, transmission impairments significantly affect data transmission efficiency. Therefore, to mitigate the impact of SOP fluctuations in a direct-detected MVM system, polarization tracking algorithms such as least-mean square (LMS), independent component analysis (ICA), Kalman filter, and extended Kalman filter (EKF) can be employed to continuously monitor and adjust the SOP of the optical signal as it propagates through the fiber [72], [73], [74], [75].

A unique three-dimensional SVM/DD system with extended reach and chromatic dispersion (CD) pre-compensation capabilities was introduced in [72]. By employing 16QAM-PAM4 polarization multiplexing over 80 km with an LMS-based MIMO DSP, they successfully achieved a transmission rate of 336 Gb/s. This result highlights the system's potential for data center applications beyond 100G and transmission distances exceeding 80 km, despite some performance degradation in the 3D SVM/DD approach caused by modulating the reference carrier.

Che et al. [69] discuss the limitations of conventional intensity modulation with direct detection beyond 100 Gb/s and introduce direct detection subsystems based on the Stokes vector receiver (SVR) to enhance spectral efficiency. The SVR, utilizing three-dimensional detection in Stokes space, enables the recovery of 2-D or 3-D modulated signals and incorporates DSP-enabled polarization demultiplexing. The proposed blind polarization demultiplexing method, combined with an LMS-based blind adaptive algorithm, simplifies

DSP by reducing two-stage polarization demultiplexing to a single-stage blind equalization, as experimentally demonstrated in the study.

Two DSP approaches for recovering intensity and interpolarization phase modulation in an SVM/DD system are explored and compared. The study introduces a novel three-stage DSP algorithm that achieves equal bit error rate (BER) performance compared to a two-stage algorithm but with significantly reduced computational complexity, accomplished by applying waveform filtering and derotation sequentially. Additionally, the proposed method improves the efficiency of polarization state tracking by a factor m, where m is the number of taps in the first filtering stage [57].

Independent component analysis (ICA) has attracted significant interest in optical communications due to its ability to separate mixed components of multidimensional stochastic data into statistically independent and uncorrelated sources. Notably, ICA-based adaptive equalizers have been developed for SVM systems to mitigate cross-polarization and inter-symbol interference (ISI). These equalizers not only de-rotate SVM constellations using ICA but also correct distorted constellation shapes. Such advanced capabilities enable promising performance even at very high symbol rates, making ICA-based equalizers a valuable tool in modern optical transmission systems [73], [76], [77].

In the domain of MIMO-OFDM transmission over few-mode fiber, ICA-based channel equalizers have proven to be highly effective. They require fewer training symbols compared to conventional methods, significantly reducing the overhead for channel equalization. Moreover, these ICA-based equalizers exhibit faster convergence, making them particularly advantageous for optical communication systems [78].

ICA-based blind equalization and phase recovery have been applied in coherent optical communication systems, demonstrating strong potential in scenarios involving QPSK and 16-QAM transmission. Notably, for 16-QAM, ICA-based equalizers outperform conventional methods such as the constant modulus algorithm (CMA). This performance advantage becomes even more significant in the presence of PMD and polarization-dependent loss (PDL), highlighting the flexibility and robustness of ICA-based equalization techniques. These results support ICA as a promising candidate to replace or complement traditional equalization methods in optical communication systems, especially in cases involving complex transmission impairments [79].

Xiang et al. [80] proposed using a Kalman filter to estimate the inverse Müller matrix in Stokes space. Additionally, an optimal tuning parameter is suggested to balance the trade-off between OSNR penalty and polarization tracking capability. As noted, although the convergence speed is enhanced, the symbol error rate (SER) improvement is only slight compared to the additional complexity introduced into the system.

In [74], a novel time-frequency domain Kalman filter structure is proposed, enabling the concurrent rather than sequential equalization of chromatic dispersion (CD) and rotation of the state of polarization (RSOP) in an SVDD system, effectively addressing the combined effects of CD and RSOP. The blind phase search (BPS) algorithm is applied to recover the original signal constellation points, though at the cost of increased overall system complexity.

Cao et al. [75] developed a sliding window-aided extended Kalman filter (SWA-EKF) method for high baud rate SVDD systems and proposed a joint impairment model that accounts for equalization-enhanced phase noise (EEPN), RSOP, and CD. For high baud rate SVDD systems, the SWA-EKF employs a sliding window structure to achieve joint equalization of EEPN, RSOP, and CD impairments. Specifically, the EKF is used to track RSOP in the time domain, compensate for CD in both frequency and time domains, and mitigate EEPN in the time domain.

Furthermore, a probabilistically shaped (PS) SVDD system with effective RSOP tracking and compensation is proposed, using a PolDemux scheme based on EKF. Unlike earlier PolDemux algorithms, the proposed method demonstrates improved performance in handling ultra-fast RSOP and is validated for both 2-parameter and 3-parameter RSOP models [81], [82].

Dissertation organization

The structure of this dissertation is as follows: the theory of MCF/MMF characterization and transmission impairments of the SDM systems are presented in chapter 2. Then, the MVM structure is analyzed in chapter 3. The performances of various DSP algorithms are discussed in chapter 4 and discussion of the fiber characterization and DSP algorithms are presented in chapter 5. Finally, conclusion and some remarks of the dissertation are presented in chapter 6.

CHARACTERIZATION OF FIBER MODAL IMPAIRMENTS USING DIRECT-DETECTION METHODS

Contribution of Authors and Co-Authors

Manuscript in chapter 2

Author: Mahmoudreza Dadras

Contributions: Implemented and tested the MD-SDM and MD-APM methods and performed the analysis. Prepared the original draft of the manuscript.

Co-Author: Ioannis Roudas

Contributions: Supervised the entire project, and provided important insight on numerical studies. Aided in the preparation of the manuscript (Review and edit).

Co-Author: Jaroslaw Kwapisz

Contributions: Provided important insight on numerical studies. Aided in the preparation of the manuscript.

Manuscript Information

Mahmoudreza Dadras, Ioannis Roudas, and Jaroslaw Kwapisz

Optics Communications

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
__X__ Published in a peer-reviewed journal

Elsevier
Vol. 523, 15 November 2022, 128735

Calculation of optimal launch modes

In [83], [84] it is shown that, in order to minimize the noise error in the estimation of the MD vector, ideally one should select an orthonormal set of Stokes vectors as launch states. Then, the errors of the MD vector components would be uncorrelated. Consider a set of N^2-1 unit vectors that form a parallelotope in the generalized Stokes space. The volume spanned by N^2-1 oblique unit vectors in the N^2-1-dimentonal Stokes space is less than unity. If the vectors are mutually orthogonal, the parallelotope becomes a hypercube, and its volume is maximized (it becomes unity). Due to the incomplete coverage of the Poincaré sphere for $N>2$, however, it is not possible to find a set of N^2-1 orthonormal Stokes vectors [10], [11]. Instead, we can compute a set of quasi-orthonormal Stokes vectors by maximizing the volume of the parallelotope

$$V=\sqrt{\det\left[\mathbf{S}\mathbf{S}^{\mathbf{T}}\right]}, \tag{2.1}$$

where det(.) denotes the determinant of a matrix and we defined the auxiliary square matrix formed from the unit vectors in Stokes space

$$\mathbf{S}=\left[\hat{s}_1,\ldots,\hat{s}_{N^2-1}\right]^T. \tag{2.2}$$

Due to the monotonicity of the logarithmic function, we adopt the objective function $\xi:=lnV$ instead of Eq. (1). The numerical optimization algorithm starts from a given set of vectors and changes their positions by small amounts successively until ξ is maximized. More specifically, we use the gradient ascent method for maximization of the objective function ξ. Assume that $\mathbf{p}$ denotes a column vector formed by the coordinates of $\hat{s}_1,\ldots,\hat{s}_{N^2-1}$. We start from an initial guess of the parameter vector $\mathbf{p_0}$ and take successive steps $\mathbf{p_k}$ in the direction of the gradient of the objective function $\nabla\xi$ until we reach a local maximum based on the recursion

$$\mathbf{p_{(k+1)}} = \mathbf{p_{(k)}} + \mu_\mathbf{k} \nabla \xi_\mathbf{k}, \tag{2.3}$$

where μ_k denotes the adaptive step size. For the analytical calculation of the components of $\nabla\xi_k$, we use Jacobi's formula for the derivative of the determinant of a matrix to obtain

$$\frac{\partial \xi}{\partial p_i} = \mathrm{Tr}[\mathbf{A}\frac{\partial \mathbf{S}}{\partial \mathbf{p_i}}], \tag{2.4}$$

where we defined the matrix $\mathbf{A} = \mathbf{S^{-1}}$ and Tr[.] represents the trace operator. Suppose that only the m-th Stokes vector depends on p_i. Then, only one element of the main diagonal of the matrix $\mathbf{A}\frac{\partial \mathbf{S}}{\partial \mathbf{p_i}}$ is non-zero. Therefore, we obtain

$$\frac{\partial \xi}{\partial p_i} = \frac{\partial \hat{s}^T}{\partial p_i} A_m, i = 1, ..., M. \tag{2.5}$$

In the following, we compute optimal sets of N^2-1 quasi-orthogonal Stokes vectors $\hat{s}_i$ that correspond to feasible combinations of propagating modes using two different gradient ascent algorithms. In the first algorithm, we parametrize the j-th unit Jones vector $|s\rangle$ using $2N-2$ hyperspherical coordinates, i.e.,

$$\begin{aligned}|s\rangle =: \big[&\cos(\phi_1), \sin(\phi_1)\cos(\phi_2)e^{i\theta_1}, \ldots, \\ &\sin(\phi_1)\cdots\sin(\phi_{N-2})\sin(\phi_{N-1})e^{i\theta_{N-1}}\big]^T.\end{aligned} \tag{2.6}$$

From this expression, we calculate the corresponding Stokes vector $\hat{s}_i$ and its derivatives with respect to ϕ_{jv} and θ_{jv}. Furthermore, we define the vector $\mathbf{p}$ that contains the coordinates ϕ_{jv} and θ_{jv} of all N^2-1 Stokes vectors. Then, we perform unconstrained optimization in a $(N^2-1)(2N-2)$ real space using the method of gradient ascent. In the second algorithm, we parametrize the j-th Jones vector $|s\rangle = (s_{jv})_{(v=1)}^{N}$ using $2N$ real parameters x_{jv} and y_{jv}. Now, the parameter vector $\mathbf{p}$ contains the coordinates x_{jv} and y_{jv} of all N^2-1 Stokes vectors. The optimization takes place in a $(N^2-1)2N$ real space, where we impose N^2-1 unit length constraints $\langle s_j|s_j\rangle$=1, $j=1,..,N^2-1$. We use the modified update rule

$$\mathbf{p_{(k+1)}} = \mathrm{proj}\left\{\mathbf{p_{(k)}} + \mu_{(\mathrm{k})}\nabla\xi_{\ \mathbf{k}}\right\}, \tag{2.7}$$

where proj(.) denotes the component of the gradient tangential to the constraints (projected gradient ascent).

Mode-dependent signal delay method (MD-SDM) measurement principle

MD-SDM is a technique developed to characterize a fiber's modal dispersion and, by extension, its MDL, using direct detection. It generalizes the polarization-dependent signal delay method (PD-SDM) used for PMD characterization in single-mode fibers [16], [11]. MD-SDM operates by sending optical pulses into the fiber, exciting specific mode combinations and the arrival times of the pulses are measured at the output. By launching controlled combinations of modes and measuring the resulting group delays, one can infer the propagation characteristics of different modes. For these measurements, linearly polarized (LP) modes are launched. LP modes, such as LP_{01} and LP_{11}, are the natural eigenmodes of MMFs and maintain stable propagation, making them ideal for measuring group delays and MD [85]. Laguerre-Gaussian (LG) beams may be used in free space to couple light into a fiber in a particular way, but inside the fiber those beams excite the conventional LP modes.

The MD-SDM consists of sending N^2-1 optical pulses using different combinations of modes through an N-mode SDM MMF and measuring the corresponding output power at the fiber output. Then, the group delays τ_{g_i} can be extracted from the output power of each mode as illustrated in Fig. 3.

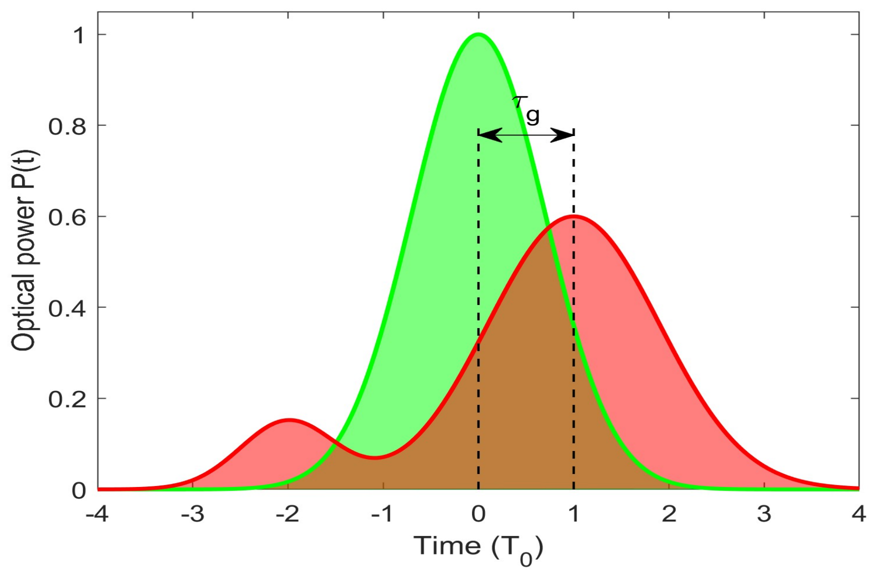

Figure 3: Input pulse (green) and output pulse (red). (Symbol: τ_g = Group delay).

The set of measured delays is then analyzed to reconstruct the fiber's MD characteristics. In practice, from these group delay measurements one can solve for the components of the fiber's MD vector. This yields the differential group delays between the principal modes of the fiber and thus quantifies the modal dispersion. In other words, the data reveals how each principal mode is delayed relative to the others.

The MD-SDM provides both the magnitude of modal dispersion (the largest delay spread between modes) and the specific mode combinations that correspond to the fastest and slowest propagation (the principal modes). This information is equivalent to identifying the fiber's principal modes and their individual DMGDs, which is important for modeling and compensating modal dispersion in SDM systems.

The MD-SDM is a purely direct-detection technique, which does not require full fiber transfer matrix or coherent receivers. Despite this simplicity, it provides results that align with more complex methods and is comparable to what one would get from a full fiber transfer matrix measurement. This makes it an attractive method for practical fiber characterization. Moreover, while the focus of MD-SDM is on measuring delay differences (MD), the overall approach can be extended to assess the MDL as well. In practice, a complementary mode-dependent average power method (MD-APM) can be used by launching the same set of mode excitations and measuring the average output power of each pulse, one can determine an MDL vector due to different attenuation coefficients in different modes. Together, the MD-SDM and the MD-APM provide a comprehensive characterization of MMFs and MCFs fiber's impairments, yielding both the MD (from group delay differences) and the MDL (from power differences) with simple measurements.

The MD-SDM can be effectively applied to MCFs to measure key transmission impairments. By selectively launching signals into different cores and analyzing the arrival times and power levels at the output, MD-SDM provides a direct way to characterize how variations between cores impact system performance [86]. This method allows precise measurement of differential group delays (core skew), which is critical for SDM systems.

In the MD-SDM formalism, launching a set of orthogonal mode combinations and measuring their delays allows computation of the fiber's MD vector, which encapsulates all the differential delays and identifies the principal modes [11]. It has been shown that launching maximally-orthogonal Stokes vectors minimizes errors in estimating the MD vector, improving SNR and the accuracy of the MD-SDM, especially as the number of modes increases [87].

Simulation of MD-SDM and MD-APM

Using the simulation setup shown in Fig. 4, we study the propagation of light through various MMFs. The waveform is generated by a continuous wave (CW) laser and modulated using a Mach-Zehnder modulator (MZM). An SLM is then employed to generate different mode combinations for propagation through the MMF. At the MMF output, optical pulses are detected by a photodetector, and the PMs and DMGDs are estimated by processing the results on a personal computer. For simplicity, we assume ideal, chirp-free, narrowband Gaussian optical pulses.

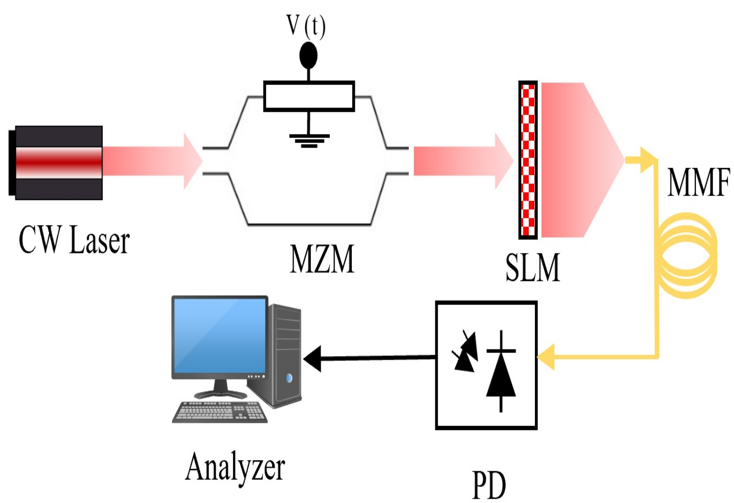

Figure 4: Simulation setup for MMF characterization using the MD-SDM (Abbreviations: CW: Continuous Wave, V(t): Bias voltage, MZM: Mach-Zehnder modulator, SLM: Spatial light modulator, MMF: Multimode fiber, PD: Photodetector).

Table 1 summarizes the parameters used in the simulation. To accurately simulate real fibers, we consider various properties of MMFs that cause different modes to exhibit different group delays, as well as frequency-dependent group delays for different supported modes. Delays associated with the polarization and spatial configurations of the LP modes are modeled by incorporating mode dispersion and intramode group delay parameters. Additionally, mode coupling is enabled to account for intragroup coupling effects. Over a span of a few hundred meters, strong coupling may occur due to the similarity of propagation constants, resulting in fully coupled modes. To capture this behavior, we introduce mean section length and mode correlation length in the simulation. Finally, to enhance accuracy, we also consider the standard deviation of the mean section length.

Table 1: Simulation parameters

Parameter	Value	Unit
Fiber length	1-20,000	m
Group refractive index	1.47	-
Attenuation	0.2-0.6	dB/km
Refractive index reference frequency	193.1×10^{12}	Hz
Core refractive index	1.45	-
Index contrast	0.02	-
Core diameter	50	µm
Cladding diameter	125	µm
Zero-dispersion slope	0.1	ps/(nm^2·km)
Zero-dispersion wavelength	1300	nm
Connector efficiency	64	%
Optical loss at connectors	2	dB
PMD coefficient	1.58×10^{-15}	$\sqrt{\text{s/m}}$
PMD correlation length	150	m
Correlation length	130	m
All modes correlation length	130	m
Effective area LP_{01}	80×10^{-12}	m^2
Mean section length	500	m
Section length deviation	5	m
Photodiode responsivity	1	A/W
Photodiode thermal noise	10^{-9}-10^{-12}	A/$\sqrt{\text{Hz}}$

Modal Dispersion

Initially, we consider the ideal case of an MMF without MDL, i.e, the fiber transfer matrix $\mathbf{U}(\omega)$ is unitary. The input group-delay operator is defined as $i\mathbf{U}^{\dagger}(\omega)\,\mathbf{U}_{\omega}(\omega)$ [11], [17], where the index ω in $\mathbf{U}_{\omega}(\omega)$ denotes differentiation with respect to the angular frequency and dagger denotes the adjoint matrix. The input group-delay operator can be expressed in terms of the identity matrix $\mathbf{I}$ and the generalized Gell-Mann matrices, $\mathbf{\Lambda} := [\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_{N^2-1}]^T$, where N is the number of modes and T denotes the transpose matrix [17]

$$i\mathbf{U}^{\dagger}(\omega)\,\mathbf{U}_{\omega}(\omega) := \tau_0(\omega)\,\mathbf{I} + \tfrac{1}{2C_N}\vec{\tau}_s(\omega)\cdot\mathbf{\Lambda}. \tag{2.8}$$

In (1), $\tau_0(\omega)$ represents the average group delay, $\vec{\tau}_s$ is the input MD vector in the generalized Stokes space, and C_N is a normalization coefficient defined as $C_N := \sqrt{N/\left[2\,(N-1)\right]}$.

The MD-SDM [17] can be used to characterize the DMGDs and input PMs of the fiber. The eigenvalues and eigenvectors of the input group-delay operator are the DMGDs and the input PMs, respectively. The group delay τ_g of an optical pulse is related to the input MD vector $\vec{\tau}_s$ and the combination of launch modes, which is represented by the Stokes vector $\hat{s}$, as follows [17]

$$\tau_g = \tau_0 + \tfrac{1}{2C_N^2}\left\langle \vec{\tau}_s \right\rangle \cdot \hat{s}, \tag{2.9}$$

where $\langle . \rangle$ denotes spectral averaging.

In (2), τ_g is defined as the time average [17]

$$\tau_g := \frac{\int\limits_{-\infty}^{\infty} tP(t)dt}{\int\limits_{-\infty}^{\infty} P(t)dt}, \tag{2.10}$$

where $P(t)$ denotes the instantaneous optical power at the fiber output. In order to measure the MD of MMFs without MDL using the MD-SDM, we employ (2.9). First, we can estimate the average group-delay τ_0 by launching N pulses corresponding to N arbitrary orthogonal launch states in Jones space. For instance, we can use the MMF eigenmodes, i.e., we can launch first all the power to the fundamental mode LP_{01} at the x-polarization; next, we can launch all the transmitted power on the mode LP_{01} at the y-polarization; and so forth. Then, the average group delay is calculated by averaging the delays of the different eigenmodes. Afterwards, we use $N^2 - 1$ optimal launch states corresponding to different spatial modes combinations into the fiber. For every launch state, we measure the power of output pulses using direct detection and calculate the group delays using (2.10). Finally, a column vector of DMGDs can be computed by (2.9) [17], and, the MD vector $\vec{\tau}_s$ can be estimated by matrix inversion [17].

Mode-dependent loss

In practice, MDL is always present in MMFs and must be characterized first because it impacts the MD measurement described in sec. 2.9. MDL is the combined result of different attenuation coefficients for different modes and of coupling among different modes due to manufacturing imperfections, fiber bends, and other mechanical deformations [17]. In [17], the authors proposed a variant of the MD-SDM, called the MD-APM, to characterize the MDL before MD.

If MDL is not negligible, the transfer matrix of the multimode fiber is not unitary and is denoted by $\mathbf{H}(\omega)$. In this case, the input group-delay operator is defined as $i\mathbf{H}_\omega(\omega)\,\mathbf{H}^{-1}(\omega)$ [17]. This is a non-Hermitian matrix with complex eigenvalues and non-orthogonal eigenvectors. To characterize MDL, indicated in [17], we can express $\mathbf{H}(\omega)$ as a product of a unitary matrix $\mathbf{U}(\omega)$ and a positive definite matrix $\mathbf{P}(\omega)$. The MDL can be determined by the ratio of the maximum and minimum of the eigenvalues of the positive semidefinite Hermitian matrix $\mathbf{P}(\omega)^2$. The latter can be written as

$$\mathbf{P}(\omega)^2 = \alpha_0(\omega)\left[\mathbf{I} + \frac{1}{2C_N}\vec{\Gamma}(\omega)\cdot\mathbf{\Lambda}\right], \tag{2.11}$$

where $\alpha_0(\omega)$ denotes the mean attenuation of the fiber and $\vec{\Gamma}(\omega)$ is the MDL vector [17]. Due to the similarity between the equation for $\mathbf{P}(\omega)^2$ and (2.8), we follow a similar procedure to the one described in sec. 2.

For MDL characterization, first, we launch N orthogonal modes in Jones space into the MMF and measure the output powers corresponding to these launched pulses. Averaging the output powers leads to the mean attenuation of the fiber $\alpha_0(\omega)$, which can be estimated by

$$\alpha_0(\omega) := \frac{1}{NP_0}\sum_{k=1}^{N} P_k. \tag{2.12}$$

$\Delta\alpha = [\Delta\alpha_1, \ldots, \Delta\alpha_{N^2-1}]^T$

Next, we launch N^2-1 mode combinations, which are independent in Stokes space, and measure the corresponding output powers P_{g_i}, $i = 1, 2, ..., N^2-1$. The matrix representation of the system of equations can be written as $\mathbf{S}\vec{\Gamma}(\omega) = \Delta\alpha$, where $\Delta\alpha = [\Delta\alpha_1, \ldots, \Delta\alpha_{N^2-1}]^T$ is a vector with elements $\Delta\alpha_i := \alpha_i - \alpha_0$ and α_i is defined as $\alpha_i := P_{g_i}/P_{\mathrm{In}}$, with P_{In} denoting the input power. Therefore, the MDL vector can be determined by solving the matrix equation

$$\vec{\Gamma}(\omega) = \mathbf{S}^{-1}\Delta\alpha, \tag{2.13}$$

where $\mathbf{S}$ denotes the matrix

$$\mathbf{S} = \left[\hat{s}_1, \ldots, \hat{s}_{N^2-1}\right]^T. \tag{2.14}$$

Subsequently, the attenuation at a distance z in the fiber can be determined using the

MDL vector and launch modes by [17]

$$\alpha(z) := \frac{\bar{P}_{Out}}{\bar{P}_{In}} = \alpha_0(\omega)\left[1 + \frac{1}{2C_N^2}\vec{\Gamma}(\omega)\cdot\hat{s}\right]. \tag{2.15}$$

Once $\mathbf{P}(\omega)$ is known, we can pre-compensate for MDL and calculate the transfer matrix of the compensated optical fiber using the method in Sec. 2. In other words, we construct new mode combinations for the pre-compensation using the formula

$$|s_k^{'}\rangle = \mathbf{P}(\omega)^{-1}|s_k\rangle, \tag{2.16}$$

where $|s_k\rangle$ and $|s_k^{'}\rangle$ denote the uncompensated and pre-compensated launch states, respectively. After calculation of the pre-compensated launch states, the corresponding mean group-delay, $\tau_0^{'}(\omega)$, and MD vector, $\vec{\tau}_s^{'}(\omega)$, can be calculated using the procedure described in sec. 2. Therefore, the compensated input group-delay operator and input group-delay operator of the fiber in the presence of MDL can be determined [17]. Subsequently, the MD vector in the presence of the MDL is complex and can be calculated by

$$\vec{\tau}_s^{'}(\omega) := C_N \mathrm{Tr}[\mathrm{i}\mathbf{H}^{-1}(\omega)\,\mathbf{H}_\omega(\omega)\cdot\mathbf{\Lambda}], \tag{2.17}$$

where Tr(.) denotes the trace of a matrix.

To summarize, first, the MD-APM can be used to characterize MDL and, subsequently, the MD-SDM can be used to characterize MD. In the joint presence of MD and MDL, initially, we characterize the MDL using the MDL formalism described in Sec. 2 and, then, we perform another set of simulations for measuring the MD vector measurement of the compensated fiber transfer matrix. Finally, we create the corresponding input group delay operator for estimating the MD vector.

Modal Crosstalk at the launch

Both the MD-SDM and MD-APM depend on the accuracy of selecting mode combinations using a mode converter, e.g., a spatial light modulator (SLM). Indeed, crosstalk is one of the most important limiting factors of the method and it can be minimized by utilizing optimal launch vectors in Stokes space [17], [88]. The phase mask implemented by the SLM by altering the voltage of the pixels can perform mode selection. The main problem of the SLM is that it may introduce crosstalk among modes [88]. It is possible to optimize the phase mask to reduce the crosstalk of SLM but even a small amount of crosstalk can affect the measurement of the input MD vector. In this section, we present a high-level model of modal crosstalk for different numbers of modes.

In order to model the modal crosstalk introduced by the mode converter without taking into account the device physics, we assume that we launch unintentionally the vector $|s_k^{'}\rangle$, instead of the desired mode combination $|s_k\rangle$ due to errors in the SLM settings [17]. In simulation, we use an abstract procedure to generate the input vectors $|s_k^{'}\rangle$. First, we generate a random vector and then find the orthonormal component $|\Delta s\rangle$ of this vector with respect to $|s_k\rangle$. The perturbed vector $|s_k^{'}\rangle$ can be expressed as $|s_k^{'}\rangle = \sqrt{1-\varepsilon}|s_k\rangle + \sqrt{\varepsilon}|\Delta s\rangle$ where ε denotes the modal crosstalk level.

Thermal noise

Thermal noise is the most significant source of noise in direct-detection optical receivers. As indicated in [17], the variance of the measured length of the MD vector due to thermal noise in the direct-detection receiver can be estimated by

$$\sigma_{N_{th}}^2 = \mathrm{Tr}\left[\mathbf{A}\mathbf{A}^T\right] \frac{N_0 C_N^4 T^3}{6R_d^2 \bar{E}^2}, \tag{2.18}$$

where $\mathbf{A} = \mathbf{S}^{-1}$. In (10), N_0, T, R_d, and $\bar{E}$ denote power spectral density (PSD) of the thermal noise, the integration time for the computation of the group delays, the responsivity of the photodiode, and the average energy of the received pulse, respectively. Then, we calculate the bound of the relative error in the MD vector by

$$B_{RE} := \frac{\sigma_{N_{th}}}{\tau_0}, \tag{2.19}$$

where $\sigma_{N_{th}}$ denotes the standard deviation of the thermal noise in the receiver.

A MULTI-DIMENSIONAL MODULATION FORMAT FOR SDM-BASED OPTICAL NETWORKS: MODE VECTOR MODULATION

Overview of mode vector modulation transceiver

Fig. 5 shows the block diagram of the entire MVM transmission system in Jones and Stokes spaces. We start by feeding data in bits b_i, to the transmitter T_x. The transmitter groups this data into blocks and transmit the Jones vectors, which is denoted by $|s_{\text{in}}\rangle$. These vectors pass through the channel, denoted by a matrix $\mathbf{H}$, which rotates the SOP of the signals. The output of the channel is the Jones vector $|s_{\text{out}}\rangle$, which is noiseless at this stage.

On the Poincaré sphere shown in Fig. 6, this operation is visualized by transmitting the symbol $|\hat{s}_{\text{in}}\rangle$ on the sphere, and due to the RSOP caused by channel $\mathbf{H}$, we should receive the symbol at a random position like $|\hat{s}_{\text{out}}\rangle$. In addition, the complex Gaussian component noise is added to the Jones vector with a certain variance. At the receiver R_x, we receive a noisy version of the Jones vector $|\vec{r}\rangle$ and convert it to Stokes vector [17], [89].

Now, because of this noise, the received Jones vector is not exactly on the surface of the Poincaré sphere and they might be above of the surface and locates at point $|\vec{r}\rangle$ in Fig. 6. This noise operates on all the symbols in the transmitted constellation, resulting in vectors that deviate from the sphere's surface. These noisy vectors are represented as Stokes vectors and denoted by $\hat{\mathbf{r}}$.

The next step in the system, following the block diagram, is the implementation of the derotator $\mathbf{W}$ in the Stokes space. It compensates for the channel-induced rotation; however, due to the noise, the corrected vectors, denoted as $\vec{s}'$, still do not lie on the Poincaré sphere. On the sphere as shown in Fig. 6, the derotated symbols shift from point $|\vec{r}\rangle$ to point $|\vec{s'}\rangle$.

A maximum-likelihood (ML) decision circuit then chooses, from all the possible transmitted symbols, the symbol that is closest in distance to the derotated received vector

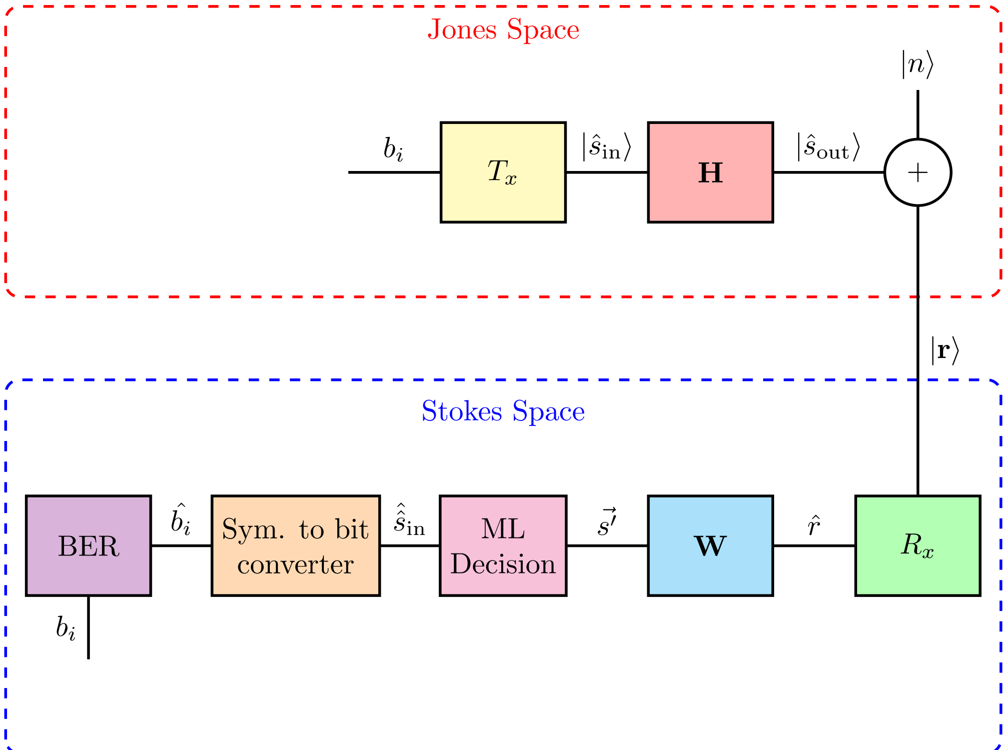

Figure 5: Block diagram of MVM transceiver in Jones and Stokes spaces. Abbreviations and symbol definitions: b_i= input bit, T_x=transmitter, $|\hat{s}_{\text{in}}\rangle$= input Jones vector, $|n\rangle$= noise in Jones space, $\mathbf{H}$= channel transfer matrix, $|\hat{s}_{\text{out}}\rangle$= output Jones vector, $|r\rangle$= output vector with noise in Jones space, R_x= receiver, $\hat{r}$= output vector with noise in Stokes space, $\mathbf{W}$= derotator, $\vec{s'}$= Stokes vector after the derotation, ML= maximum likelihood, $\hat{\hat{s}}_{\text{in}}$= estimated Stokes vector, $\hat{b}_i$= estimated input bit, BER= bit error rate.

$\mathbf{S}'$ based on a maximum-likelihood criterion. The estimated symbol, $\hat{s}_{\text{in}}$, may not be exactly the originally transmitted symbol s_{in}, so we denote it as $\hat{\hat{s}}_{\text{in}}$.

Finally, we map these estimated symbols back to bits using a symbol-to-bit converter, creating estimated bits $\hat{b}_i$. These estimated bits are compared to the original transmitted bits b_i, and the bit error rate (BER) is computed as the number of errors between the transmitted and received bits over the total number of bits.

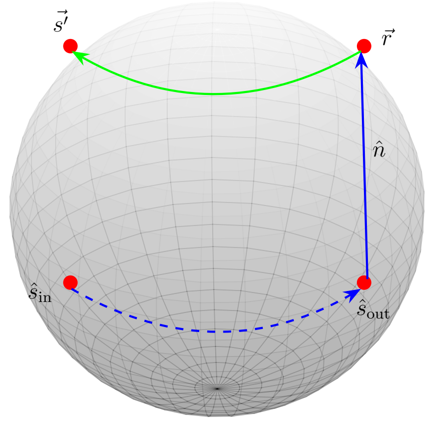

Figure 6: Rotation and derotation of symbols on the Poincaré sphere.

Polarization tracking modeling in the generalized Stokes Space

Time-varying channel propagation in the MVM-DD system can be characterized by the Muller matrix. The inverse Muller matrix along the fiber must be identified in order to apply an equalizer to track the state of polarization.

Rotational form in Stokes and generalized Stokes spaces

The rotational forms in Stokes space can be denoted by [89]

$$\hat{t} = \mathbf{R}\hat{s}, \tag{3.1}$$

where $\hat{t}$, $\hat{s}$, and $\mathbf{R}$ denote the output Stokes vector, input Stokes vector, and Muller matrix, respectively. The connection between Stokes and Jones spaces can be defined by [89]

$$\mathbf{R}\vec{\sigma} = \mathbf{U}^{\dagger}\vec{\sigma}\mathbf{U}, \tag{3.2}$$

where $\vec{\sigma}$ is Pauli spin vector in Stokes space.

Consider an N-dimensional Jones space and a $D = (N^2 - 1)$-dimensional Stokes space. Rotations within Jones space are given by $|t\rangle = \mathbf{U}\,|s\rangle$ for a unitary Jones matrix $\mathbf{U}$, while rotations within Stokes space are given by $\hat{t} = \mathbf{R}\hat{s}$ for a Mueller matrix $\mathbf{R}$. We can calculate the Mueller matrix $\mathbf{R}$ from the Jones matrix $\mathbf{U}$:

$$\mathbf{R}_{ij} = \frac{1}{2}\,\mathrm{tr}\left[\mathbf{U}^{\dagger}\lambda_i\mathbf{U}\lambda_j\right] \tag{3.3}$$

where λ_i, λ_j denote the generalized Gell-Mann matrices.

[Proof of (3.3)] Letting $|s\rangle$ and $\hat{s}$ denote the same state of polarization in Jones and Stokes space respectively, we will perform the rotations $|t\rangle = \mathbf{U}\,|s\rangle$ and $\hat{t} = \mathbf{R}\hat{s}$ and seek to relate the resulting terms. We begin by recalling that the components of the Stokes vector $\hat{s}$ are given by

$$\hat{s}_k = C_N \langle s|\lambda_k|s\rangle\,, \quad C_N = \sqrt{\frac{N}{2(N-1)}}, \tag{3.4}$$

where λ_k is a generalized Gell-Mann matrix. Using that $|t\rangle = \mathbf{U}\,|s\rangle$ and converting to Stokes space yields

$$\hat{t}_i = C_N \langle t|\lambda_i|t\rangle = C_N \langle s|\mathbf{U}^{\dagger}\lambda_i\mathbf{U}|s\rangle\,, \tag{3.5}$$

while using $\hat{t} = \mathbf{R}\hat{s}$ yields

$$\hat{t}_i = \sum_k \mathbf{R}_{ik}\hat{s}_k = \sqrt{\frac{N}{2(N-1)}} \sum_k \mathbf{R}_{ik} \langle s|\lambda_k|s\rangle$$
$$= C_N \left\langle s \middle| \sum_k \mathbf{R}_{ik}\lambda_k \middle| s \right\rangle . \tag{3.6}$$

Equating these two expressions for $\hat{t}_i$ yields that

$$\langle s|\mathbf{U}^\dagger \lambda_i \mathbf{U}|s\rangle = \left\langle s \middle| \sum_k \mathbf{R}_{ik}\lambda_k \middle| s \right\rangle . \tag{3.7}$$

As this equation holds for arbitrary $|s\rangle$, we arrive at the matrix equation

$$\mathbf{U}^\dagger \lambda_i \mathbf{U} = \sum_k \mathbf{R}_{ik}\lambda_k, \tag{3.8}$$

which yields

$$\mathbf{U}^\dagger \lambda_i \mathbf{U}\lambda_j = \sum_k \mathbf{R}_{ik}\lambda_k\lambda_j \tag{3.9}$$

upon right-multiplication by λ_j. Taking the trace of both sides, while noting that $\mathrm{tr}\,[\lambda_k\lambda_j] = 2\delta_{kj}$, gives that

$$\mathrm{tr}\left[\mathbf{U}^\dagger \lambda_i \mathbf{U}\lambda_j\right] = \sum_k \mathbf{R}_{ik}\,\mathrm{tr}\,[\lambda_k\lambda_j]$$
$$= \sum_k \mathbf{R}_{ik} 2\delta_{kj} = 2\mathbf{R}_{ij} \tag{3.10}$$

and thus

$$\mathbf{R}_{ij} = \frac{1}{2}\,\mathrm{tr}\left[\mathbf{U}^\dagger \lambda_i \mathbf{U}\lambda_j\right] \tag{3.11}$$

as desired.

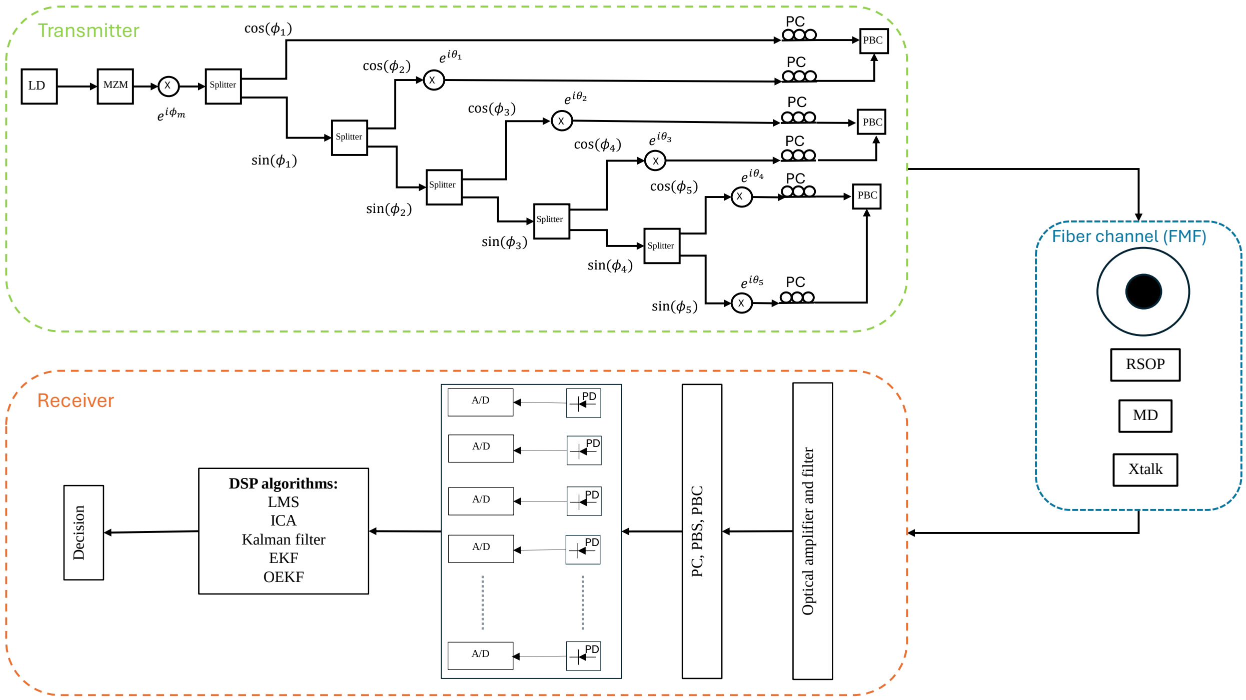

Figure 7: Block diagram of the MVM transmission system. Symbols: LD=laser diode, MZM=Mach-Zehnder modulator, PC=polarization controller, PBC=polarization beam combiner, PBS=polarization beam splitter, RSOP=rotation of state of polarization, MD=modal dispersion, PD=photodetector, A/D= analog-to-digital converter.

Simulation setup

Fig. 7 represents the block diagram of the MVM optical transmission system under study in this paper. The details of design of the MVM transceiver are given in [71]. A brief overview of the transmission system based on the MVM is presented in this section.

At the transmitter, optical pulses are generated using a laser diode, followed by a Mach-Zehnder modulator and a phase modulator. The pulses are then split into N parallel branches with varying amplitudes using of electro-optic splitters. Then individual branches use an array of phase modulators, introducing phase differences between the vector components.

The polarization controllers and polarization beam combiners are used to combine the different optical paths and get orthogonal SOPs, which are subsequently launched into different fiber modes. For the purpose of this paper, the transmitter generates $(6,64)$-MVM signals as discussed in [71] with the following features. The pulses are shaped using a root-raised cosine (RRC) filter and transmitted at different symbol rates over few-mode fibers (FMFs) with 3 spatial modes and 2 polarization modes ($N = 6$). Our simulations take into account all the most important linear impairments, i.e., RSOP, MD, and crosstalk. We used Monte Carlo simulations with various levels of crosstalk but most of the simulations have been done with an average crosstalk value of -30 dB/km.

At the receiver, signals are optically pre-amplified and, then, different spatial and polarization components of the each beranch have to be separated. This is done by means of mode demultiplexers and polarization beam splitters. Further polarization controllers and polarization beam splitters/combiner are used to recombine the polarization components. An array of $2N^2 - N$ photodiodes subsequently measures the power of these components which is followed by analog-to-digital converters. Finally, maximum likelihood estimation is used in Stokes space, and the BER is tested.

To compare the performance of different equalization techniques in the presence of various impairments like RSOP, MD, and crosstlak, we simulate a (6,64)-MVM and present results of comparing equalization methods such as LMS, ICA, EKF, and OEKF.

Table 2 shows the parameters that were used in the simulation.

Table 2: Simulation parameters

Parameters	Value	Unit
DMD	0,3.02072e-17, 3.02072e-17	ps/km
CD	2.16914e-05, 2.21603e-05, 2.21603e-05	ps/(nm.km)
CD slope	67.09935, 68.51562, 68.51562	ps/(nm^2.km)
Fiber attenuation	0.16401, 0.15887, 0.15887	dB/km
Fiber length	0.5-40	km
Symbol rate	14-130	GBd
Spatial and polarization degrees of freedom	6	-
Constellation points	32-1024	-
Crosstalk	-50 to -10	dB/km

ADAPTIVE EQUALIZATION METHODS FOR STOKES AND MODE VECTOR MODULATION

Contribution of Authors and Co-Authors

Manuscript in chapter 4

Author: Mahmoudreza Dadras

Contributions: Implemented and tested all ICA, Kalman filter, and EKF. Developed and tested the OEKF method, and prepared the manuscript.

Co-Author: Ioannis Roudas

Contributions: Supervised the entire project, and provided important insight on numerical studies. Aided in the preparation of the manuscript (Review and edit).

Manuscript Information

Mahmoudreza Dadras, and Ioannis Roudas

IEEE Photonics Journal

Status of Manuscript:
_X_ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Institute of Electrical and Electronics Engineering (IEEE)

Principle and algorithms of various DSP methods

Multidimensional modulation formats in communication systems require robust equalization techniques to mitigate the effects of channel impairments. Four prominent approaches include LMS, ICA, the Kalman filter, and the extended Kalman filter (EKF). LMS is an adaptive filtering algorithm that iteratively adjusts filter coefficients to minimize the mean square error between transmitted and received signals. ICA aims to separate statistically independent sources from the received signal. In contrast, the Kalman filter and EKF are recursive algorithms used to estimate the state of a dynamic system affected by noise and uncertainties. In the following sections, we provide a detailed comparative analysis of these four DSP algorithms within the MVM-DD framework for derotating the rotation of the state of polarization.

Least-Mean Squares (LMS)

Fig. 8 shows the flowchart of the LMS algorithm. This algorithm iteratively adjusts the weights of a linear filter to minimize the mean squared error between the predicted and desired outputs [90].

The process initiates with the initialization of two essential parameters: a weight vector $\mathbf{w}$ and a step size μ. These provide the foundation for the iterative adjustments that follow. Next, the algorithm predicts the output $\hat{y}_i$ based on the current weight vector $\mathbf{w}_i$ and the input vector $\mathbf{x}_i$. This prediction represents the algorithm's estimate of the output at the current iteration. It then iteratively refines the model by updating the weight vector $\mathbf{w}_{i+1}$. This update is driven by the difference between the desired output $\mathbf{d}_i$ and the predicted output $\hat{y}_i$, which guides the modification of the filter's parameters.

The weight update is then followed by a decision on whether the iterations should continue or if the algorithm should be terminated. This decision is based on predefined conditions, such as reaching a maximum iteration limit or achieving a satisfactory level of convergence. At the same time, the error e_i is updated, representing the difference between the desired and predicted outputs. This step determines how the weights should be adjusted in the next iteration. The adaptive nature of the algorithm is further emphasized by the update of the step size μ_{i+1}, which depends on the current step size μ_i, the norm of the input vector $\|\mathbf{x}_i\|$, and a small positive constant ϵ that helps control the convergence behavior. These steps are repeated until the algorithm meets the predefined termination conditions. Since it is iterative and its parameters update adaptively, the LMS algorithm efficiently minimizes prediction errors, making it a versatile tool for adaptive filtering applications

LMS algorithm The LMS algorithm is used to update the weights of a filter to minimize the mean square error between an estimated and a desired output. Parameters for the LMS are initialized, and the weights of the filter are assigned small random values, $\mathbf{W}$.

In each cycle of the main loop, the algorithm processes the available measurements. It takes as input the current vector, $\mathbf{X}_i$, and specifies as output the desired output $\mathbf{D}_i$, which corresponds to a reference signal. The estimated output, $\mathbf{Y}_i$, is then computed by taking the dot product of the filter weights $\mathbf{W}$ and the input vector $\mathbf{X}_i$.

The error, e_i, is computed as the difference between the desired output $\mathbf{D}_i$ and the corresponding estimated output $\mathbf{Y}_i$. The core concept of the algorithm is that the filter weights are updated according to the LMS update rule. This update is performed in small steps, μ, to gradually adapt the filter weights based on the computed error.

Due to its iterative nature, the LMS algorithm is well-suited for real-time applications where adjustments to the filtering process are necessary as new data becomes available. By minimizing the mean square error, the algorithm enhances its predictive capabilities and adapts to changing conditions, making it valuable in a wide range of signal processing and adaptive filtering scenarios.

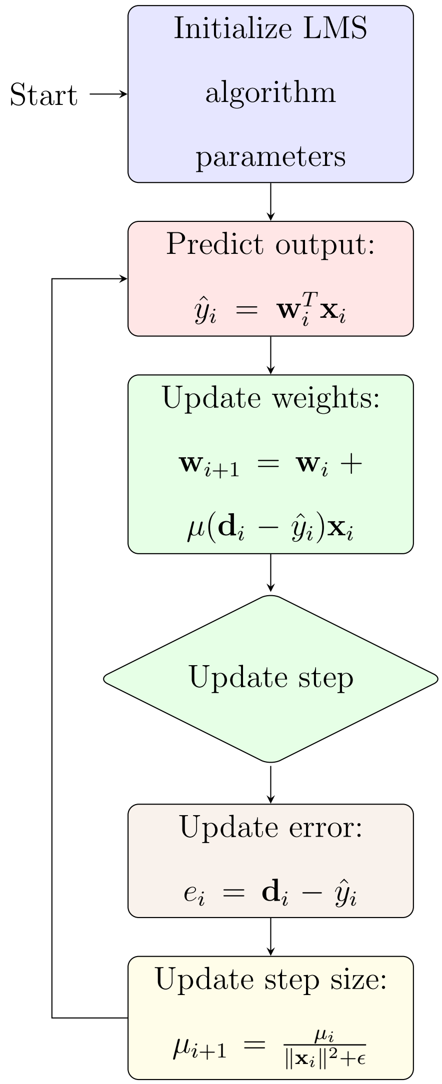

Figure 8: Flowchart of the least-mean square (LMS) algorithm.

Algorithm 1. Least Mean Square Algorithm

Initialize LMS parameters

Initialize filter weights $\mathbf{W}$ with small random values

while measurement available **do**

Input current observation vector $\mathbf{X}_i$

Desired output $\mathbf{D}_i$ (ground truth or reference signal)

Calculate the estimated output:

$\mathbf{Y}_i = \mathbf{W}^T\mathbf{X}_i$

Compute the error:

$e_i = \mathbf{D}_i - \mathbf{Y}_i$

Update filter weights:

$\mathbf{W}_{i+1} = \mathbf{W}_i + \mu \cdot e_i \cdot \mathbf{X}_i$ {Update with step size μ}

end while=0

LMS computational complexity The LMS algorithm performs different computational steps that add up to its overall complexity. From table 3, the process initializes initial parameters and variables with the constant complexity of $O(1)$. Prediction of output $\hat{y}_i$ through a matrix-vector multiplication is responsible for developing a complexity of $O(N^2)$, where N is the number of samples. The algorithm performs a weight update $\mathbf{w}_{i+1}$ for every prediction error, which has a complexity of $O(N^2)$. The error update e_i, calculating the error as the difference between the desired and actual outputs, is an $O(N^2)$ operation. Finally, updating the step size has a complexity of $O(N^2)$, and the overall iteration complexity of the LMS algorithm turns out to be $O(N^2)$.

Table 3: Complexity analysis of LMS algorithm

Operation/Step	**Complexity**
Initialization	$O(1)$
Predict output: $\hat{y}_i = \mathbf{w}_i^T \mathbf{x}_i$	$O(N^2)$
Update weights: $\mathbf{w}_{i+1} = \mathbf{w}_i + \mu(\mathbf{d}_i - \hat{y}_i)\mathbf{x}_i$	$O(N^2)$
Update error: $e_i = \mathbf{d}_i - \hat{y}_i$	$O(N^2)$
Update step size: $\mu_{i+1} = \frac{\mu_i}{\|\mathbf{x}_i\|^2+\epsilon}$	$O(N^2)$
Overall Iteration Complexity	$O(N^2)$

Independent component analysis (ICA)

Independent component analysis (ICA) is the blind source separation of mixed signals into their original underlying source signals. It assumes that the sources generate uncorrelated signals that are statistically independent of one another. ICA consists of two major tasks: first, whitening the data to simplify the separation process; and second, rotating the components to maximize their statistical independence.

Whitening is a crucial step that transforms the data into an orthogonal form with a known covariance matrix. In the context of ICA, whitening simplifies the computational task of finding independent components. The ICA algorithm then applies a fixed-point iteration to rotate the pre-whitened components, aiming to achieve maximum independence according to measures of non-Gaussianity.

It starts by setting up the parameters of the ICA algorithm. Centering is essential to ensure that there are no biases within the data. This step is performed by subtracting the mean vector from each data point. It is a critical preprocessing step in transforming the data to obtain components that are uncorrelated and have unit variance. This is achieved by applying a whitening matrix to the centered data, ensuring that the resulting features are statistically independent.

Fig. 9 contains a flowchart illustrating an update step, which iteratively refines the estimate of the mixing matrix $\mathbf{W}$ to better separate the independent components. The update step includes a learning rate term, μ, and applies the function $g(\mathbf{Y}_i)$ to the whitened data matrix.

An orthogonalization process is then performed on the mixing matrix to ensure that it remains orthogonal after each update. This step promotes independence and prevents re-mixing of the extracted components during learning. The algorithm then returns to the centering step, establishing an iterative cycle aimed at refining the mixing matrix for optimal independent component separation.

The ICA algorithm is based on the linear mixing model given by [76], [77]

$$\mathbf{X} = \mathbf{A}\mathbf{S} + N, \tag{4.1}$$

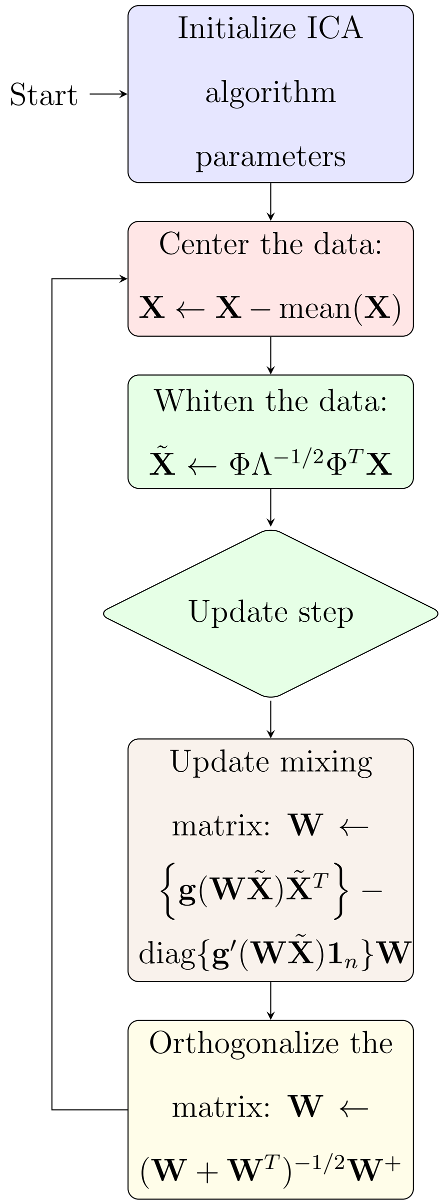

Figure 9: Flowchart of the independent component analysis (ICA) algorithm.

where $\mathbf{X}$ is the received signal vector, $\mathbf{A}$ is the mixing matrix, $\mathbf{S}$ is the source signal vector, and N is the noise.

We can express the relationship between the source signal vector $\mathbf{S}$ and the demixing matrix $\mathbf{W}$ as

$$\mathbf{S}' = \mathbf{W}\mathbf{X}. \tag{4.2}$$

For the whitening process, the received signal vector is transformed as follows [73]

$$\tilde{\mathbf{X}} = \Phi\Lambda^{-1/2}\Phi^T\mathbf{X}, \tag{4.3}$$

where Φ and Λ are the eigenvector and eigenvalue matrices of the covariance matrix of $\mathbf{X}$, respectively.

The ICA adaptation process consists of several key steps [77], [76], [73]. First, centering is performed by subtracting the mean vector of the output data to ensure it is centered around zero. Next, zero-phase whitening is applied to transform the raw data into a dataset with unit variance and uncorrelated components. Following this, a fixed-point iteration method is employed to compute a unitary unmixing matrix $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)^T$ that maximizes non-Gaussianity. The matrix $\mathbf{W}$ is then updated using the

$$\mathbf{W}^+ \leftarrow \left\{\mathbf{g}(\mathbf{W}\tilde{\mathbf{X}})\tilde{\mathbf{X}}^T\right\} - \mathrm{diag}\{\mathbf{g}'(\mathbf{W}\tilde{\mathbf{X}})\mathbf{1}_n\}\mathbf{W}, \tag{4.4}$$

where $\mathbf{g}$ and $\mathbf{g}'$ represent a non-Gaussianity measure and its derivative, respectively, and $\mathbf{1}_n$ is an $n \times 1$ vector of ones. Finally, orthogonalization is applied to refine $\mathbf{W}$ using the transformation:

$$\mathbf{W} \leftarrow (\mathbf{W} + \mathbf{W}^T)^{-1/2}\mathbf{W}^+, \tag{4.5}$$

where $\mathbf{W}^{+}$ denotes the pseudo-inverse of $\mathbf{W}$.

ICA algorithm The ICA algorithm can be used to derotate the SOP by separating mixed polarization components into statistically independent sources. First, the algorithm centers the input data $\mathbf{X}$, which aligns the signal for further processing. Then, it applies zero-phase whitening to decorrelate the components and equalize variances, effectively rotating the polarization states into a new orthonormal basis. Fixed-point iteration iteratively refines the unmixing matrix $\mathbf{W}$ to maximize non-Gaussianity, which helps separate independent polarizations. The orthogonalization step ensures that $\mathbf{W}$ remains different at each iteration, which improves the orthogonality of SOPs. Through these steps, the algorithm adaptively rotates and unmixes the polarization components, compensating for polarization mixing caused by transmission impairments in the fiber.

ICA computational complexity Table 4 provides computational complexities for different steps of the ICA algorithm. In the initialization step, there is a complexity of $O(1)$, which implies constant computational effort that does not depend on the size of the input. Centering, which involves subtracting the mean vector from the input signal vector $\mathbf{X}$, has a complexity of $O(N)$.

Next, the zero-phase whitening procedure rotates the data orthogonal to the axes such that the resulting components are uncorrelated with unit variances.

The complexity of fixed-point iteration depends on the inner operations and the number of iterations performed during the execution of the ICA algorithm. It varies with different convergence behaviors. Additionally, the use of specific non-Gaussianity measures $\mathbf{g}$ and $\mathbf{g}'$ affects the complexity of the update operation for $\mathbf{W}$.

Orthogonalization ensures that the unmixing matrix remains orthogonal and has a complexity of $O(N^3)$. Therefore, the overall computational complexity of the ICA algorithm is $O(N^3)$.

Algorithm 2. Independent Component Analysis algorithm

Initialize ICA parameters

Initialize mixing matrix $\mathbf{A}$ and estimate of sources $\mathbf{S}$

while measurement available **do**

Centering: Ensure the output data is centered around zero by subtracting its mean vector $\mathbf{X}_{\text{centered}} = \mathbf{X} - \text{mean}(\mathbf{X})$

Zero Phase Whitening: Transform the raw data to a dataset with uncorrelated components and unit variances $\tilde{\mathbf{X}} = \Phi\Lambda^{-1/2}\Phi^T\mathbf{X}_{\text{centered}}$

Fixed Point Iteration: Find a unitary unmixing matrix $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)^T$ that maximizes non-Gaussianity.

Update $\mathbf{W}$: Update the step size $\mathbf{W}$ $\mathbf{W}^+ \leftarrow \left\{\mathbf{g}(\mathbf{W}\tilde{\mathbf{X}})\tilde{\mathbf{X}}^T\right\} - \text{diag}\{\mathbf{g}'(\mathbf{W}\tilde{\mathbf{X}})\mathbf{1}_n\}\mathbf{W}$

Orthogonalization $\mathbf{W} \leftarrow (\mathbf{W} + \mathbf{W}^T)^{-1/2}\mathbf{W}^+$

end while=0

Table 4: Computational Complexity Analysis of ICA Algorithm

Operation/Step	**Complexity**
Initialization	$O(1)$
Centering: Subtract the mean vector from data	$O(N)$
Whitening: $\tilde{\mathbf{X}} = \Phi\Lambda^{-1/2}\Phi^T\mathbf{X}$	$O(N^2)$
Fixed-Point Iteration: Compute unitary unmixing matrix $\mathbf{W}$ to maximize non-Gaussianity	$O(N^2)$
Update $\mathbf{W}$: $\mathbf{W}^+ \leftarrow \left\{\mathbf{g}(\mathbf{W}\tilde{\mathbf{X}})\tilde{\mathbf{X}}^T\right\} - \mathrm{diag}\{\mathbf{g}'(\mathbf{W}\tilde{\mathbf{X}})\mathbf{1}_n\}\mathbf{W}$	$O(N^2)$
Orthogonalization: $\mathbf{W} \leftarrow (\mathbf{W} + \mathbf{W}^T)^{-1/2}\mathbf{W}^+$	$O(N^3)$
Overall Iteration Complexity	$O(N^3)$

Kalman filter

The Kalman filter is a mathematical algorithm that processes a sequence of measurements to infer information about the previous states of a system and predict its current state. As shown in the flowchart in Fig. 10, the fundumental concept of the Kalman filter to estimate the system state using a prediction model and then refine it based on incoming measurements [91], [92].

The two primary stages of the Kalman filter are prediction and update. In the prediction step, the filter forecasts the state of the system at the next time step based on the current state and any control inputs, using a mathematical model of the system. In the update step, the filter refines this prediction using new measurement data. The updated estimate is computed as a weighted average of the prediction and the measurement, assigning greater weight to the measurement when it is more accurate [91]. The Kalman filter method also utilizes matrices that model the system dynamics and measurement noise in a form that allows the optimal weights to be determined for the prediction and update steps. As more measurements become available, the filter continually updates its estimate of the system state, resulting in progressively improved accuracy over time [91].

When applied in combination with Stokes vector modulation, the Kalman filter estimates the optical signal's SOP and updates it by considering fiber polarization effects. It enhances the received signal quality by using signal measurements to improve the SOP estimation and dynamically adjusts it to maintain a stable state of polarization [81], [92].

In practical implementations, a combination of hardware components—such as frequency and polarization tracking loops—and software algorithms performing the Kalman filter computations can be used for polarization-state tracking applications.

In order to track the SOP in SVM/DD and MVM/DD systems, the inverse Müller matrix in Stokes space must be estimated. A Müller matrix represents the polarization properties of a system and relates the input and output polarization states at both ends of an optical link. Its inverse may be used to model polarization-dependent effects within the system [80].

In the Kalman filter technique, a state-space model is used to monitor the temporal evolution of the SOP and to compute the inverse Müller matrix. Using the Kalman filter algorithm, the inverse Müller matrix can be estimated from input and output polarization data, with the state-space model developed by deriving the physical relationships between these polarization states [92].

This recursive approach allows for highly accurate estimates of the inverse Müller matrix, even in the presence of measurement noise or other uncertainties. Furthermore, the Kalman filter algorithm can be adapted to accommodate varying system conditions, such as rapid changes in SOP or the presence of polarization-dependent impairments [74], [75].

The basic structure of a Kalman filter in SVDD or MVM-DD systems includes several key components [92]: the state model, measurement model, state estimator, error covariance matrix, and gain matrix.

The state model captures the temporal evolution of the polarization state of a signal and defines the system dynamics. Given the current state, it enables the prediction of the system's future state. The measurement model describes how the observed measurements relate to the true state of polarization and is used to estimate the system state based on incoming data.

At the core of the Kalman filter is the state estimator, which computes the estimated system state. It combines information from the state model and the measurement model to provide an accurate estimate of the system state based on current measurements.

The error covariance matrix quantifies the uncertainty of the state estimate. It is updated continuously using the results of state prediction and measurement comparison, allowing the filter to adapt to varying conditions.

Finally, the gain matrix updates the state estimate by weighing the relative confidence in the predicted state and the new measurement. Since the gain matrix is influenced by the evolving error covariance matrix, the Kalman filter dynamically adapts to changing system conditions.

<u>Kalman filter algorithm</u> This section presents the implementation of an iterative Kalman filter that can be used to estimate the state of a system based on a series of measurements performed at the MVM/DD receiver. The algorithm first predicts the next state using a mathematical model of the system and then corrects this prediction using actual measurements.

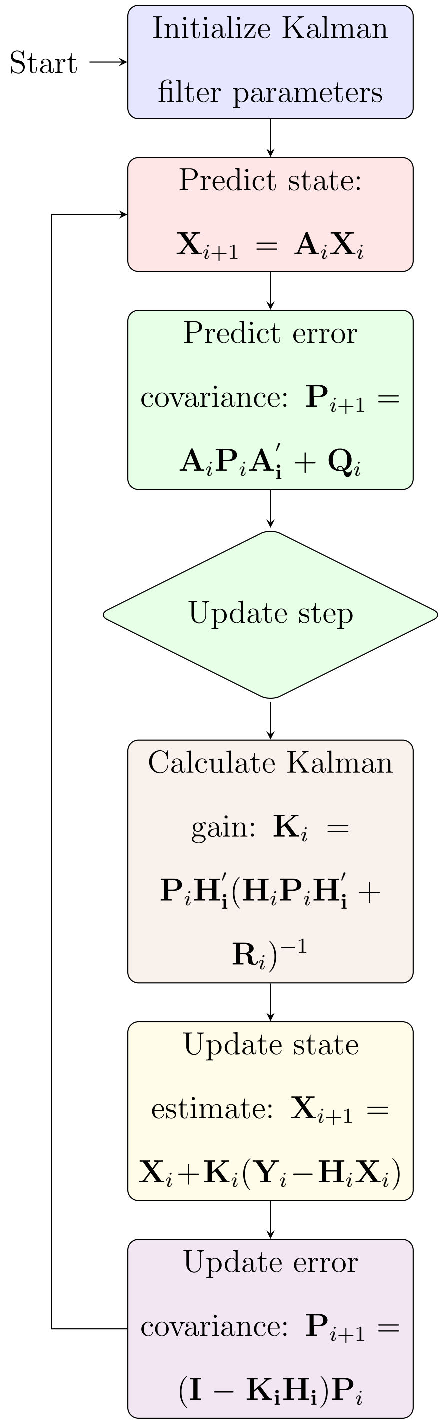

Figure 10: Flowchart of the Kalman filter algorithm

First, we need to predict the next state based on the previous state $\mathbf{X}_i$. Therefore, the next state can be calculated by [91]

$$\mathbf{X}_{i+1} = \mathbf{A}_i \mathbf{X}_i, \tag{4.6}$$

where $\mathbf{A}_i$ and $\mathbf{X}_i$ are the transition matrix and the current state matrix, respectively. In the absence of any external inputs or measurements, the transition matrix describes the linear relationship between the current state estimate and the subsequent state estimate. It captures the dynamics of the system, including how the state evolves over time and the influence of any external inputs.

Next, the uncertainty or error in the state estimate is predicted using the error covariance matrix of the next state. This matrix plays a critical role in the Kalman filter, as it helps balance the influence of new measurements with the predicted state, thereby producing an optimal estimate of the system state. The error covariance of the next state can be denoted by [91]

$$\mathbf{P}_{i+1} = \mathbf{A}_i \mathbf{P}_i \mathbf{A}_\mathbf{i}^{'} + \mathbf{Q}_i, \tag{4.7}$$

where $\mathbf{P}_i$ and $\mathbf{Q}_i$ are the current error covariance and the process noise, respectively. Note that the process noise represents the variability or uncertainty in the system model that the Kalman filter does not explicitly account for. This noise is modeled as a zero-mean random variable with an associated covariance matrix, which is used to adjust the error covariance matrix and incorporate the uncertainty in the state estimation.

In the next step, the Kalman gain must be defined to optimally combine the predicted state estimate and the measurement, yielding the best estimate of the system state. The Kalman gain determines how much weight is assigned to the measurement versus the predicted state during the correction step.

When the measurement is highly accurate, the Kalman gain assigns greater weight to the measurement and less to the predicted state. Conversely, when the measurement is less reliable, more weight is given to the predicted state and less to the measurement. This dynamic weighting allows the filter to account for both measurement uncertainty and model uncertainty, producing an optimal state estimate.

The Kalman gain minimizes the estimation error by adjusting the balance between the predicted state and the new measurement, ultimately leading to a more accurate and reliable estimate of the system state.

To compute the Kalman gain, the measurement noise covariance matrix is required. This matrix models the uncertainty associated with the measurements by capturing the statistical properties of the measurement noise. The Kalman gain can therefore be determined by [91]

$$\mathbf{K}_i = \mathbf{P}_i\mathbf{H}_\mathbf{i}^{\prime}(\mathbf{H}_i\mathbf{P}_i\mathbf{H}_\mathbf{i}^{\prime} + \mathbf{R}_i)^{-1}, \tag{4.8}$$

where $\mathbf{R}_i$ and $\mathbf{H}_i$ are the measurement noise covariance and the measurement matrices, respectively. A very important part of the Kalman filter is the measurement matrix, commonly referred to as the observation matrix. This matrix maps the state variables into the measurement space. It is used to compare the expected measurements with the actual observed data and serves as a link between the predicted state estimate and the real measurements.

Now, we need to update the predicted state estimate using the actual measurement $\mathbf{Y}_i$, which is given by [91]

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{K}_i(\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i), \tag{4.9}$$

and also the error covariance can be updated by [91]

$$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K_i H_i})\mathbf{P}_i. \tag{4.10}$$

Algorithm 3. Kalman Filter Algorithm

Initialize Kalman filter parameters

Initialize state estimate $\mathbf{X}$ and error covariance $\mathbf{P}$

while measurement available **do**

Predict state:

$\mathbf{X}_{i+1} = \mathbf{A}\mathbf{X}_i$

$\mathbf{P}_{i+1} = \mathbf{A}\mathbf{P}_i\mathbf{A}' + \mathbf{Q}$

Update step:

Calculate Kalman gain:

$\mathbf{K}_i = \mathbf{P}_i\mathbf{H}'(\mathbf{H}\mathbf{P}_i\mathbf{H}' + \mathbf{R})^{-1}$

Update state estimate:

$\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{K}_i(\mathbf{Y} - \mathbf{H}\mathbf{X}_i)$

Update error covariance:

$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\mathbf{H})\mathbf{P}_i$

end while=0

Kalman filter computational complexity Table 5 presents a complexity analysis of the Kalman filter equations, detailing the computational costs associated with each step. The initialization step has a constant complexity of $O(1)$, reflecting its efficiency regardless of the input size.

The prediction state equation $\mathbf{X}_{i+1} = \mathbf{A}_i\mathbf{X}_i$ involves matrix-vector multiplication and has a complexity of $O(N^2)$. The error covariance prediction step $\mathbf{P}_{i+1} = \mathbf{A}_i\mathbf{P}_i\mathbf{A}'_i + \mathbf{Q}_i$ requires matrix-matrix multiplication and addition, resulting in a complexity of $O(N^3)$.

Calculating the Kalman gain $\mathbf{K}_i = \mathbf{P}_i\mathbf{H}'_i(\mathbf{H}_i\mathbf{P}_i\mathbf{H}'_i + \mathbf{R}_i)^{-1}$ involves matrix multiplications, additions, and a matrix inversion whose complexity depends on the specific algorithm used.

Updating the state estimate $\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{K}_i(\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i)$ includes matrix-vector multiplication and has a complexity of $O(N^2)$. Updating the error covariance $\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i$ involves matrix multiplications and subtraction, resulting in a complexity of $O(N^3)$.

The overall iteration complexity of the Kalman filter equations is dominated by the $O(N^3)$ terms, highlighting the computational cost associated with each iteration of the filtering algorithm.

Extended Kalman filter (EKF)

The typical Kalman filter assumes a linear relationship between the measurements and the state variables, using linear equations to model the system dynamics. However, in many applications—such as tracking the state of polarization—nonlinear relationships exist between the state variables and the measurements due to polarization-related impairments. These nonlinearities make it difficult for linear models to accurately capture the system's dynamics. As a result, the standard Kalman filter may fail to provide precise estimates in such scenarios and, in some cases, may even diverge [75], [93]. The nonlinear system dynamics and measurement models are linearized by the EKF utilizing the Jacobian matrix to resolve this issue. The Jacobian matrix depicts the linearized approximation of the measurement models and nonlinear system dynamics based on the most recent estimate of the state. In order to generate updated state estimates that are more accurate for nonlinear systems, the EKF applies this linearized approximation in the Kalman filter algorithm's prediction and

Table 5: Complexity analysis of Kalman filter algorithm

Operation/Step	**Complexity**
Initialization	$O(1)$
Predict state: $\mathbf{X}_{i+1} = \mathbf{A}_i\mathbf{X}_i$	$O(N^2)$
Predict error covariance: $\mathbf{P}_{i+1} = \mathbf{A}_i\mathbf{P}_i\mathbf{A}'_i + \mathbf{Q}_i$	$O(N^3)$
Calculate Kalman gain: $\mathbf{K}_i = \mathbf{P}_i\mathbf{H}'_i(\mathbf{H}_i\mathbf{P}_i\mathbf{H}'_i + \mathbf{R}_i)^{-1}$	$O(N^3)$
Update state estimate: $\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{K}_i(\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i)$	$O(N^2)$
Update error covariance:$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i$	$O(N^3)$
Overall Iteration Complexity	$O(N^3)$

update steps [91].

In order to implement the EKF, as shown in Fig. 11, we need to define the state transition function $f(X_n)$ and the measurement function $h(X_n)$. Therefore, the predicted next state is given by [82], [91], [93]

$$\mathbf{X}_{i+1} = f(\mathbf{X_n}). \tag{4.11}$$

The error covariance of the next state is [82], [91], [93]

$$\mathbf{P}_{i+1} = \frac{\partial f}{\partial x}\mathbf{P}_i(\frac{\partial f}{\partial x})' + Q, \tag{4.12}$$

where the $\frac{\partial f}{\partial x}$ denotes the Jacobian matrix of the state transition matrix and $(\frac{\partial f}{\partial x})'$ is the

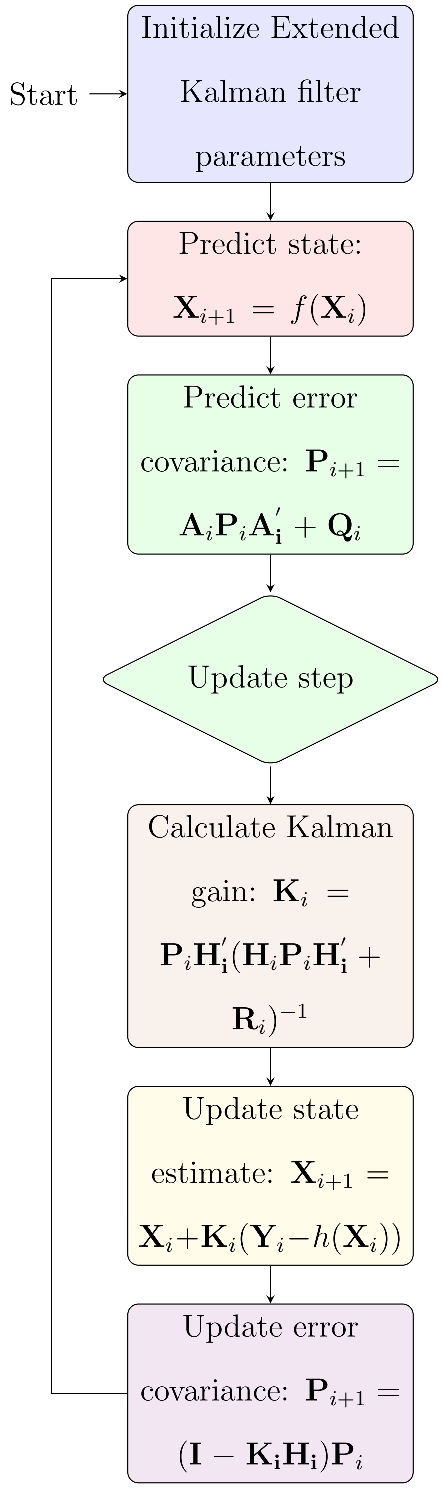

Figure 11: Flowchart of the extended Kalman filter algorithm

transpose of the Jacobian matrix. Then, the Kalman gain can be calculated using [82], [91], [93]

$$\mathbf{K}_i = \mathbf{P}_i(\frac{\partial h}{\partial x})'(\frac{\partial h}{\partial x}\mathbf{P}_i(\frac{\partial h}{\partial x})' + \mathbf{R})^{-1}, \tag{4.13}$$

where the $\frac{\partial h}{\partial x}$ is the Jacobian matrix of the covariance [82], [91], [93]. Finally, the updated state estimate using actual measurement is [82], [91], [93]

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{K}_i(\mathbf{Y}_i - h(\mathbf{X}_i)), \tag{4.14}$$

and also the updated error covariance can be denoted by [82], [91], [93]

$$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\frac{\partial h}{\partial x})\mathbf{P}_i. \tag{4.15}$$

<u>Extended Kalman filter algorithm</u> The EKF is an iterative estimation algorithm designed for state estimation in nonlinear systems. It begins by initializing the algorithm-specific parameters and setting up the initial state estimate $\mathbf{X}$ along with its corresponding error covariance matrix $\mathbf{P}$. The algorithm iteratively processes available measurements through a prediction-update loop. In the prediction step, the state prediction $\hat{\mathbf{X}}_{i+1}$ is obtained by applying the nonlinear system dynamics function f to the current state estimate $\hat{\mathbf{X}}_i$. The Jacobian matrix of the system dynamics function, $\mathbf{A}$, is calculated to linearize the nonlinear dynamics, and the error covariance for the next state, $\mathbf{P}_{i+1}$, is predicted using the linearized dynamics and the process noise covariance matrix $\mathbf{Q}$.

Moving to the update step, the Kalman gain K is computed based on the linearized measurement model. This involves calculating the Jacobian matrix of the measurement model, $\mathbf{H}$, at the predicted state $\hat{\mathbf{X}}_{i+1}$. The state estimate is then updated using the Kalman gain and the difference between the actual measurement $\mathbf{Y}$ and the predicted measurement from the measurement model. Simultaneously, the error covariance is updated using the Kalman gain to account for the reduction in uncertainty after incorporating the measurement.

These steps are iteratively repeated for each available measurement, refining both the state estimate and the error covariance. The algorithm adapts to nonlinear systems by linearizing the dynamics and measurement models at each iteration. In essence, the EKF algorithm leverages predictions from nonlinear models and updates from real measurements to iteratively estimate the state of a dynamic system while accounting for uncertainties in both the system dynamics and the observations.

EKF computational complexity The computational complexity of the EKF is dictated by the matrix operations involved in the prediction and update steps. Table 6 summarizes the complexity of each step. The initialization step has a constant complexity of $O(1)$.

The error covariance prediction step, $\mathbf{P}_{i+1} = \mathbf{A}_i\mathbf{P}_i\mathbf{A}_i' + \mathbf{Q}$, involves matrix multiplications and additions. Since $\mathbf{P}$ is an $N \times N$ matrix, the dominant operation in this step is the multiplication of $\mathbf{A}_i$ (an $N \times N$ matrix) with $\mathbf{P}_i$ and its transpose, resulting in a complexity of $O(N^3)$.

Similarly, the Kalman gain computation requires inverting the matrix $\mathbf{H}_i\mathbf{P}_i\mathbf{H}_i' + \mathbf{R}$, which is also of size $N \times N$. Matrix inversion has a computational complexity of $O(N^3)$, making this one of the most computationally expensive steps in the EKF.

The state update step, $\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{K}_i(\mathbf{Y}_i - h(\mathbf{X}_i))$, consists primarily of matrix-vector multiplications. Since $\mathbf{X}$ is an $N \times 1$ vector and $\mathbf{K}$ is an $N \times N$ matrix, the computational cost is $O(N^2)$. In contrast, the error covariance update, $\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i$, requires matrix multiplications between $N \times N$ matrices, leading to a complexity of $O(N^3)$.

Since the most computationally intensive operations in the EKF involve matrix multiplications and matrix inversions, the overall iteration complexity is dominated by

Algorithm 4. Extended Kalman Filter Algorithm

Initialize Extended Kalman filter parameters

Initialize state estimate $\mathbf{X}$ and error covariance $\mathbf{P}$

while measurement available **do**

Predict state:

$\hat{\mathbf{X}}_{i+1} = f(\hat{\mathbf{X}}_i)$ {Non-linear prediction using system dynamics}

$A = \frac{\partial f}{\partial \hat{\mathbf{X}}}|_{\hat{\mathbf{X}}=\hat{\mathbf{X}}_i}$ {Jacobian matrix of the system dynamics}

$\mathbf{P}_{i+1} = \mathbf{A}\mathbf{P}_i\mathbf{A}' + \mathbf{Q}$

Update step:

Calculate Kalman gain:

$\mathbf{H} = \frac{\partial h}{\partial \hat{\mathbf{X}}}|_{\hat{\mathbf{X}}=\hat{\mathbf{X}}_{i+1}}$ {Jacobian matrix of the measurement model}

$\mathbf{K}_i = \mathbf{P}_i\mathbf{H}'(\mathbf{H}\mathbf{P}_i\mathbf{H}' + R)^{-1}$

Update state estimate:

$\mathbf{X}_{i+1} = \hat{\mathbf{X}}_{i+1} + \mathbf{K}_i(\mathbf{Y}_i - h(\hat{\mathbf{X}}_i))$ {Non-linear update using measurement model}

Update error covariance:

$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\mathbf{H})\mathbf{P}_i$

end while=0

the $O(N^3)$ terms. This implies that, for large state dimensions, the EKF can become computationally demanding—particularly in real-time applications where fast, repeated updates are required.

Optimized extended Kalman filter (OEKF)

In this section, we discuss the optimization of the cost function to find the optimal state vector $f(X)$ that minimizes the error between the received and transmitted signals using the

Table 6: Complexity analysis of EKF

Operation/Step	**Complexity**
Initialization	$O(1)$
Error Covariance Prediction: $\mathbf{P}_{i+1} = \mathbf{A}_i \mathbf{P}_i \mathbf{A}_i' + \mathbf{Q}$	$O(N^3)$
Kalman Gain Calculation: $\mathbf{K}_i = \mathbf{P}_i \mathbf{H}_i' (\mathbf{H}_i \mathbf{P}_i \mathbf{H}_i' + \mathbf{R})^{-1}$	$O(N^3)$
State Update: $\mathbf{X}_{i+1} = \mathbf{X}_{i+1} + \mathbf{K}_i (\mathbf{Y}_i - h(\mathbf{X}_{i+1}))$	$O(N^2)$
Error Covariance Update: $\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i \mathbf{H}_i) \mathbf{P}_i$	$O(N^3)$
Overall Iteration Complexity	$O(N^3)$

extended Kalman filter algorithm. The cost function $\mathbf{J}(X_i)$ is defined as:

$$\mathbf{J}(\mathbf{X}) = \sum_{i=1}^{N} \|\mathbf{y}_i - \mathbf{H}_i \mathbf{X}_i\|^2 \tag{4.16}$$

To minimize this cost function, we use the gradient descent algorithm. The update rule for gradient descent is:

$$\mathbf{X}_{i+1} = \mathbf{X}_i - \alpha \nabla \mathbf{J}(\mathbf{X}_i), \tag{4.17}$$

where $\mathbf{X}_i$ is the current estimate of the state vector, α is the learning rate, and $\nabla \mathbf{J}(\mathbf{X}_i)$ is the gradient of the cost function with respect to $\mathbf{X}$. Therefore, the gradient can be computed as

$$\nabla J(\mathbf{X}) = -2\sum_{i=1}^{N} \mathbf{H}_i^T(\mathbf{y}_i - \mathbf{H}_i\mathbf{X}_i). \tag{4.18}$$

Now, we can update the state vector iteratively using gradient descent until convergence:

$$\mathbf{X}_{i+1} = \mathbf{X}_i + 2\alpha\sum_{i=1}^{N} \mathbf{H}_i^T(\mathbf{y}_i - \mathbf{H}_i\mathbf{X}_i). \tag{4.19}$$

The flowchart shown in Fig. 12 depicts the workflow of the optimized extended Kalman filter (OEKF) algorithm, designed for state estimation in nonlinear systems. The process starts with initializing the filter parameters, including the state vector $\mathbf{X}$, error covariance matrix $\mathbf{P}$, process noise covariance $\mathbf{Q}$, and measurement noise covariance $\mathbf{R}$. The algorithm predicts the next state using the system dynamics $\mathbf{X}_{i+1} = f(\mathbf{X}_i)$, while updating the error covariance matrix $\mathbf{P}_{i+1}$ based on the process model and noise characteristics. This prediction serves as the basis for subsequent corrections.

During the update phase, the Kalman gain $\mathbf{K}_i$ is computed to optimally merge the predicted state with incoming measurements, reducing estimation errors. The state estimate and error covariance are updated iteratively in this phase. If the error fails to converge, the algorithm calculates the cost function and its gradient to fine-tune the parameters $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{P}$ using optimization techniques. This optimization loop allows the filter parameters to adapt dynamically to minimize estimation errors.

The process continues until the error converges to a minimum, at which point the algorithm stops. By integrating optimization into the EKF framework, this approach enhances its performance, particularly in systems with complex or nonlinear dynamics.

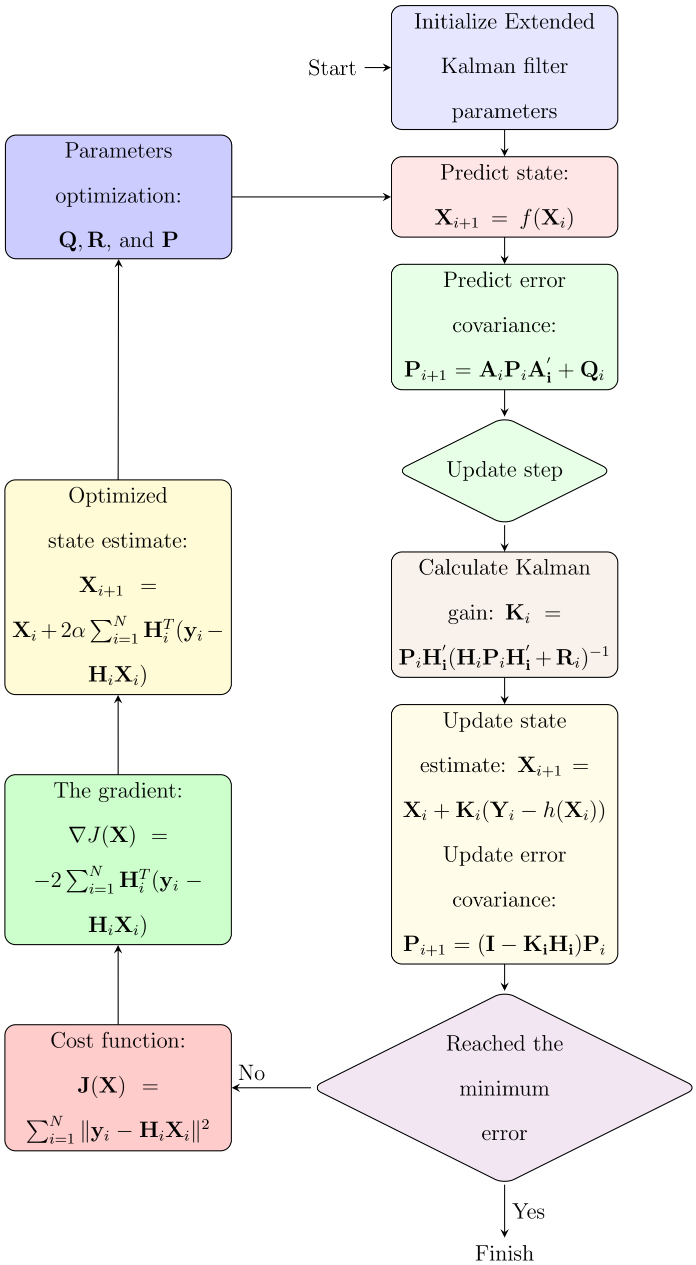

Figure 12: Flowchart of the optimized extended Kalman filter (OEKF) algorithm.

OEKF algorithm The OEKF algorithm enhances the traditional EKF by incorporating optimization steps that dynamically adjust filter parameters for improved estimation accuracy. The algorithm begins with the initialization of state estimates, error covariance, and noise covariance matrices. It follows the standard EKF procedure, consisting of a prediction step—where the state estimate is propagated through a nonlinear transition model—and an update step, where new measurements refine the state estimate using the Kalman gain. The error covariance is also updated based on the Jacobians of the state transition and measurement models.

A key distinguishing feature of OEKF is its optimization step, which minimizes a cost function representing the error between predicted and actual measurements. The algorithm computes the gradient of this cost function and updates the state estimate iteratively using a gradient-based approach. Additionally, it fine-tunes the $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{P}$ matrices to ensure optimal parameter selection. This adaptive approach enhances the filter's robustness and accuracy, particularly in dynamic and uncertain environments. The algorithm iterates through prediction, update, and optimization steps until convergence is achieved, ensuring refined state estimation over time.

<u>OEKF computational complexity</u> The OEKF consists of several computational steps, each contributing to the overall complexity of the algorithm. Table 7 presents the complexity analysis of these operations. The initialization step has a complexity of $O(1)$, as it involves setting the initial state estimates, covariance matrices, and noise parameters. The state prediction step, $\hat{\mathbf{X}}_{i+1} = f(\hat{\mathbf{X}}_i)$, depends on the system model but is typically an $O(N)$ operation, assuming a moderately complex function.

The prediction step for the error covariance matrix, $\mathbf{P}_{i+1} = \mathbf{A}_i\mathbf{P}_i\mathbf{A}_i' + \mathbf{Q}$, is computationally intensive as it requires matrix multiplications. Since $\mathbf{P}$ is an $N \times N$ matrix, multiplying it by $\mathbf{A}_i$ results in a complexity of $O(N^3)$. Similarly, computing the Kalman gain, $\mathbf{K}_i = \mathbf{P}_i\mathbf{H}_i'(\mathbf{H}_i\mathbf{P}_i\mathbf{H}_i' + \mathbf{R})^{-1}$, involves a matrix inversion, which is also $O(N^3)$. The state update step, $\mathbf{X}_{i+1} = \hat{\mathbf{X}}_i + \mathbf{K}_i(\mathbf{Y}_i - h(\hat{\mathbf{X}}_i))$, consists primarily of matrix-vector multiplications, leading to a lower complexity of $O(N^2)$. The error covariance update, $\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i$, involves additional matrix multiplications, resulting in $O(N^3)$ complexity.

The optimization step in OEKF adds further computations compared to the standard EKF. The cost function evaluation, $\mathbf{J}(\mathbf{X}) = \sum_{i=1}^{N} \|\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i\|^2$, and gradient computation, $\nabla\mathbf{J}(\mathbf{X}) = -2\sum_{i=1}^{N} \mathbf{H}_i^T(\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i)$, both have a complexity of $O(N^2)$. Updating the parameters, $\mathbf{X}_{i+1} = \mathbf{X}_i + 2\alpha\sum_{i=1}^{N} \mathbf{H}_i^T(\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i)$, is also $O(N^2)$. However, the optimization of $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{P}$ requires iterative computations with a complexity of $O(N^3)$. Since the most computationally demanding steps involve $O(N^3)$ operations, the overall iteration complexity remains at $O(N^3)$. This suggests that while OEKF provides improved accuracy through parameter optimization, it comes at a higher computational cost—making it more suitable for applications where estimation accuracy is prioritized over computational efficiency.

Overall, the computational complexity of different equalization algorithms varies based on the operations involved. The LMS algorithm primarily consists of matrix-vector multiplications and weight updates, leading to a dominant iteration complexity of $O(N^2)$, which is relatively efficient for adaptive filtering applications.

ICA exhibits higher computational complexity compared to LMS due to its iterative nature and matrix operations. The major computational steps of ICA include centering ($O(N)$), whitening ($O(N^2)$), fixed-point iteration for computing the unmixing matrix ($O(N^2)$), iterative weight update ($O(N^2)$), and orthogonalization ($O(N^3)$). Among these, the orthogonalization step dominates the complexity, leading to an overall iteration complexity of $O(N^3)$. As a result, ICA has more complexity than LMS but remains a viable option for applications requiring source separation in high-dimensional modulation formats.

In contrast, the Kalman filter and its variants, such as the EKF and OEKF, involve even more computationally demanding operations. The standard Kalman filter includes matrix-vector and matrix-matrix multiplications, as well as matrix inversions. The prediction state update and state estimation steps exhibit $O(N^2)$ complexity, while the error covariance prediction and Kalman gain computation involve $O(N^3)$ complexity due to matrix-matrix multiplications and inversions. Thus, the overall iteration complexity of the Kalman filter is dominated by $O(N^3)$. The EKF and OEKF introduce additional optimization steps, further increasing computational costs while improving estimation accuracy. These higher complexities make Kalman-based approaches more computationally intensive than LMS, particularly in high-dimensional applications.

Comparing these algorithms, LMS remains the most computationally efficient, followed by ICA, which requires additional matrix computations and iterative updates. Kalman filtering and its variants demand the highest computational resources, making them the most expensive in terms of complexity. The choice of algorithm depends on the trade-off between computational efficiency and accuracy requirements in DSP applications.

Table 7: Complexity analysis of OEKF algorithm

Operation/Step	**Complexity**
Initialization	$O(1)$
Predict state: $\hat{\mathbf{X}}_{i+1} = f(\hat{\mathbf{X}}_i)$	$O(N)$
Predict error covariance: $\mathbf{P}_{i+1} = \mathbf{A}_i \mathbf{P}_i \mathbf{A}_i' + \mathbf{Q}$	$O(N^3)$
Compute Kalman gain: $\mathbf{K}_i = \mathbf{P}_i \mathbf{H}_i' (\mathbf{H}_i \mathbf{P}_i \mathbf{H}_i' + \mathbf{R})^{-1}$	$O(N^3)$
Update state: $\mathbf{X}_{i+1} = \hat{\mathbf{X}}_i + \mathbf{K}_i(\mathbf{Y}_i - h(\hat{\mathbf{X}}_i))$	$O(N^2)$
Update covariance: $\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i \mathbf{H}_i)\mathbf{P}_{i+1}$	$O(N^3)$
Evaluate cost: $\mathbf{J}(\mathbf{X}) = \sum_{i=1}^{N} \lVert \mathbf{Y}_i - \mathbf{H}_i \mathbf{X}_i \rVert^2$	$O(N^2)$
Compute gradient: $\nabla \mathbf{J}(\mathbf{X}) = -2 \sum_{i=1}^{N} \mathbf{H}_i^T (\mathbf{Y}_i - \mathbf{H}_i \mathbf{X}_i)$	$O(N^2)$
Update parameters: $\mathbf{X}_{i+1} = \mathbf{X}_i + 2\alpha \sum_{i=1}^{N} \mathbf{H}_i^T (\mathbf{Y}_i - \mathbf{H}_i \mathbf{X}_i)$	$O(N^2)$
Optimize $\mathbf{Q}, \mathbf{R}, \mathbf{P}$	$O(N^3)$
Overall Complexity	$O(N^3)$

Algorithm 5. Optimized Extended Kalman Filter (OEKF) Algorithm

Initialize Extended Kalman filter parameters: state estimate $\mathbf{X}_0$, error covariance $\mathbf{P}_0$, and noise covariances $\mathbf{Q}$ and $\mathbf{R}$.

while measurements are available **do**

Predict state:

$\hat{\mathbf{X}}_{i+1} = f(\hat{\mathbf{X}}_i)$ {State transition model}

Predict error covariance:

$\mathbf{P}_{i+1} = \mathbf{A}_i\mathbf{P}_i\mathbf{A}_i^{'} + \mathbf{Q}$ {Jacobian $\mathbf{A}_i = \frac{\partial f}{\partial \mathbf{X}}$}

Update Step:

$\mathbf{K}_i = \mathbf{P}_i\mathbf{H}_i^{'}(\mathbf{H}_i\mathbf{P}_i\mathbf{H}_i^{'} + \mathbf{R})^{-1}$

Update state estimate:

$\mathbf{X}_{i+1} = \hat{\mathbf{X}}_i + \mathbf{K}_i(\mathbf{Y}_i - h(\hat{\mathbf{X}}_i))$ {Measurement model $h(\cdot)$}

Update error covariance:

$\mathbf{P}_{i+1} = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i$ {Jacobian $\mathbf{H}_i = \frac{\partial h}{\partial \mathbf{X}}$}

Optimization Step:

Evaluate cost function:

$\mathbf{J}(\mathbf{X}) = \sum_{i=1}^{N} \|\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i\|^2$

Compute gradient:

$\nabla\mathbf{J}(\mathbf{X}) = -2\sum_{i=1}^{N} \mathbf{H}_i^T(\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i)$

Update parameters:

$\mathbf{X}_{i+1} = \mathbf{X}_i + 2\alpha\sum_{i=1}^{N} \mathbf{H}_i^T(\mathbf{Y}_i - \mathbf{H}_i\mathbf{X}_i)$

Optimize $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{P}$.

if parameters are optimized **then**

Proceed to the next prediction step.

else

Recalculate parameters and continue optimization.

RESULTS AND DISCUSSION

Results and discussion

Maximizing the volume of the parallelotope

To improve the accuracy of using the MD-SDM and MD-APM, we applied a maximization procedure for computing optimal launch modes. Fig. 13 illustrates volume reduction as a function of the number of modes N. The different curves in Fig. 13 show the results of the gradient ascent algorithm starting from various initial conditions. Specifically, the blue, green, and red curves are obtained using the unconstrained gradient ascent algorithm initialized with vectors proposed by Yang and Nolan [94], vectors selected from mutually unbiased bases (MUBs) [95], and vectors based on symmetric, informationally-complete, positive operator-valued measures (SIC-POVM) [96], respectively.

The points represent the results obtained using the projected gradient ascent algorithm. The superior performance of the SIC-POVM vectors can be attributed to their high degree of symmetry—specifically, their equal pairwise angular separations. This symmetry allows the optimization algorithm to converge more rapidly toward a set of quasi-orthogonal vectors in Stokes space.

One way to quantify the obliqueness of a vector set is by calculating the inner products of each pair of vectors in the set, forming the Gram matrix. As a specific example, Fig. 14 (a) and (b) show density plots of the absolute value of the Gram matrix before and after optimization, respectively. In this example, we use vectors selected from MUBs as the initial conditions. The optimization results in a set of nearly perpendicular vectors, with the Gram matrix approaching a unit matrix.

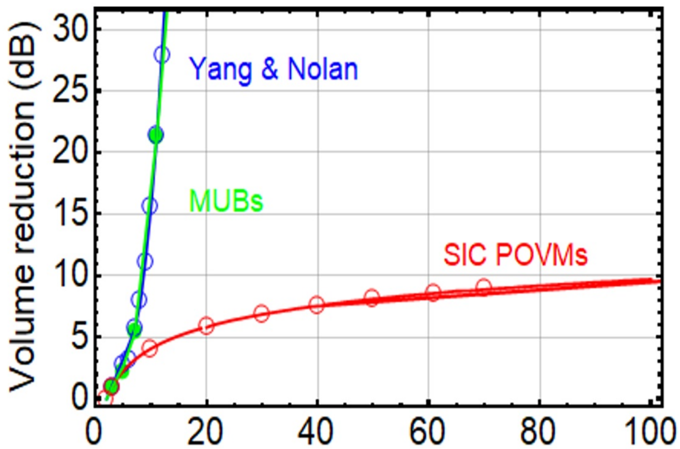

Figure 13: Parallelotope volume decrease in dB vs. the number of modes. (Initial conditions: Red: SIC-POVM vectors; Green: Vectors from MUBs; Blue: Yang and Nolan's vectors. Methods: Points: unconstrained optimization; Lines: projected gradient ascent algorithm).

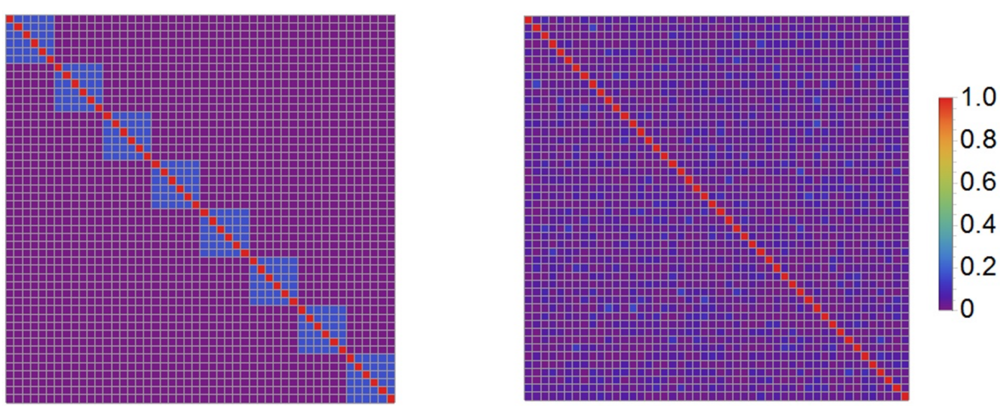

Figure 14: Density plots of the Gram matrix for N=7 vector from MUBs (a) before and (b) after optimization.

Fig. 15 shows the number of iterations required for convergence by the unconstrained gradient ascent algorithm as a function of the number of fiber modes.

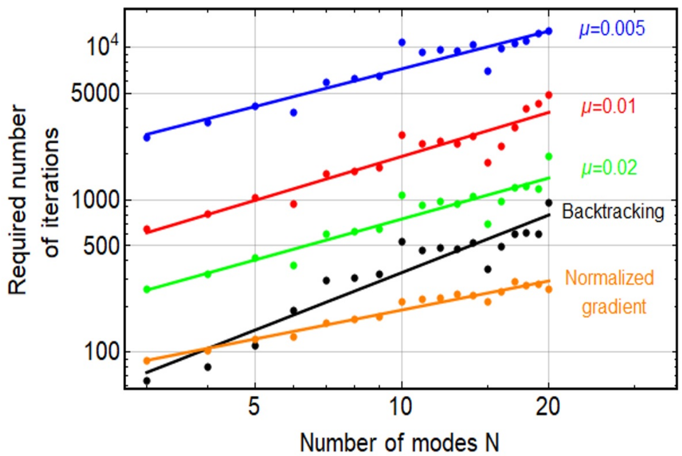

Figure 15: Comparison of required iterations to achieve convergence vs the number of modes using SIC-POVM vectors as initial conditions (Condition: ϵ_ζ=10^{-2}), μ: Adaptive step size.

Finally, Fig. 16 shows that using gradient ascent to maximize the volume of the parallelotope results in a set of Stokes vectors that nearly—albeit not completely—minimize the noise-based cost function described in [84]. Therefore, volume maximization serves as a practical alternative optimization approach to noise minimization. The key advantage of this method is a slight gain in computational efficiency.

Fiber characterization using the MD-SDM and the MD-APM

In this section, we present simulation results for the characterization of different fibers using the MD-SDM and the MD-APM. The accuracy in estimating the norm of the MD vector is used as a figure of merit for evaluating the performance of the MD-SDM.

Fig. 17 compares the magnitude of the MD vector as a function of wavelength, derived from the transfer matrix (red) and estimated using the MD-SDM (blue), in the absence of MDL, for a 500 m long 6-mode fiber. We observe a small discrepancy between the MD

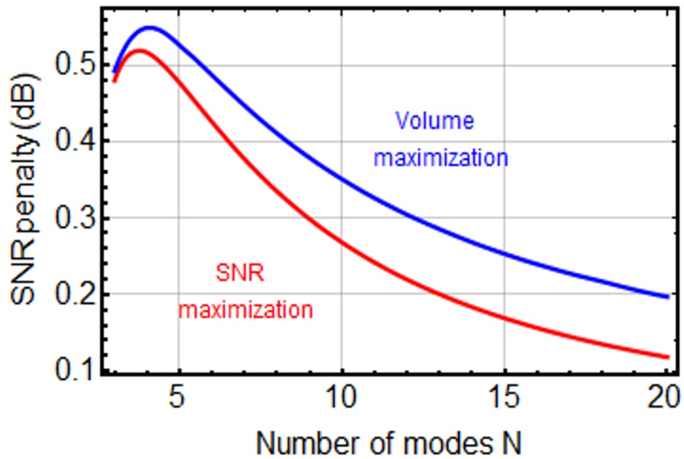

Figure 16: SNR penalty (dB) vs number of fiber modes for SIC POVM vectors as an initial condition (Symbols: Blue: set of vectors given by volume maximization; Red: given by noise minimization).

vector estimates obtained by the two methods. Specifically, the largest difference is 1.22 ps, occurring at 1537.5 nm within the C-band.

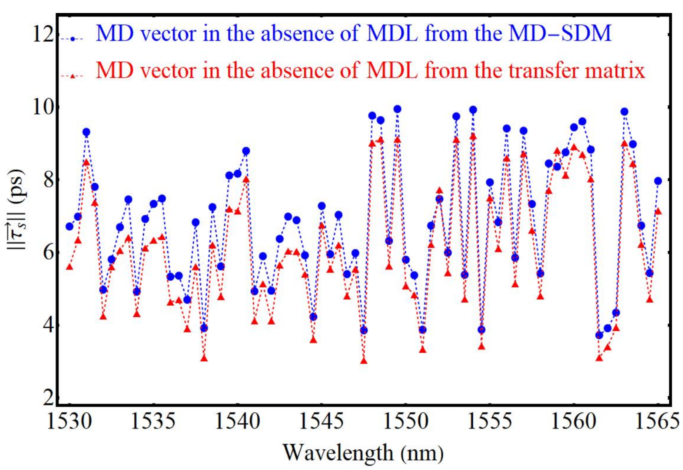

Figure 17: MD vector estimation in the absence of the MDL using the MD-SDM and the fiber transfer matrix.

When the fiber has MDL, its transfer matrix is no longer unitary, and the MD vector must be calculated using the process described in Chapter 2.

Fig. 18 shows the MD vector estimation using the same process, despite the presence of MDL. As expected, the MD estimation error is larger in this case, since the effect of MDL is ignored. The maximum error is 3.41 ps and occurs at 1552.5 nm.

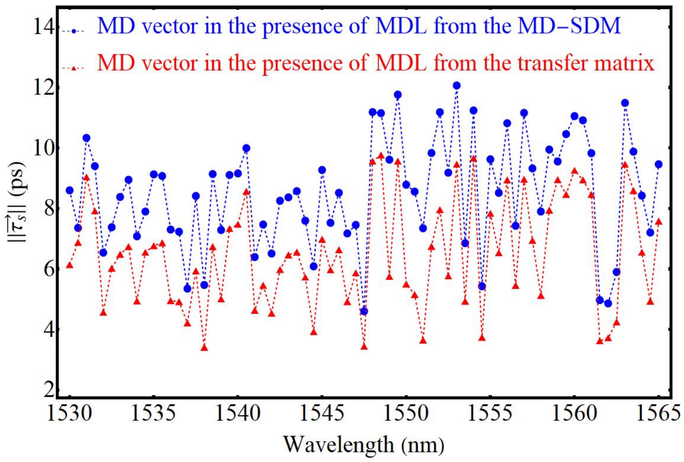

Figure 18: MD vector estimation in the presence of the MDL using the MD-SDM and transfer matrix of the fiber.

MDL plays a fundamental role in limiting signal propagation in SDM fibers. As mentioned above, the presence of MDL affects the characterization of the MD vector, and therefore, accurate MDL estimation is vital when using the MD-SDM method.

In addition, the transfer matrix of the fiber can be calculated using the input and output PMs provided by the simulation tool. The transfer matrix is given by $\mathbf{H}(\omega) = \mathbf{PM}_{Out}(\omega) \cdot \mathbf{PM}_{In}(\omega)^{-1}$, and the MDL can be estimated from the ratio of the largest to the smallest eigenvalue of $\mathbf{P}(\omega)^2 = \mathbf{H}(\omega)^{\dagger} \cdot \mathbf{H}(\omega)$ [17].

Fig. 19 shows the MDL and the effect of receiver thermal noise on the MDL measurement in the C-band, as estimated from both the transfer matrix and the MD-SDM. At 1536 nm, a 0.1 dB deviation is observed between the estimated and actual MDL of a 500 m 6-mode fiber.

Mode coupling in MMFs is another factor that has an impact on the performance of a transmission system. Indeed, by studying the mode coupling, we estimate the accuracy of the MD-SDM at different coupling regimes.

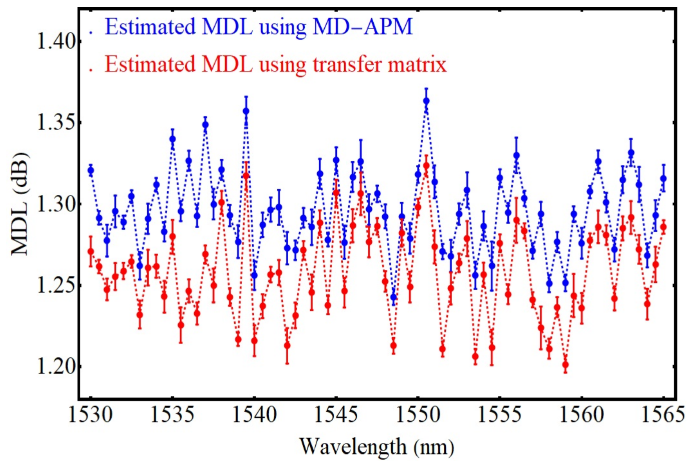

Figure 19: Comparison of the estimated MDL by the MD-APM and the actual MDL derived by the transfer matrix.

Kahn et al. [97] expressed the strength of mode coupling as a function of the correlation length and the length of the fiber. We study the impact of weak and strong coupling on the accuracy of the MD-SDM and the MD-APM. We define the coupling parameter as $\delta_c := L_c/L_s$, where L_c and L_s denote the correlation length and the fiber length, respectively. Using this definition, when δ_c is close to one, we are in the weak-coupling regime. When δ_c approaches zero, the correlation length is much shorter than the system length, and we are in the strong-coupling regime.

Fig. 20 shows the impact of different coupling regimes on the MD vector estimation using the MD-SDM. Strong coupling reduces the separation of group delays, and it can compensate for MD vector estimation errors. For example, for a 1 km 6-mode fiber, the penalty for the MD vector estimation is less than -25 dB.

Fig. 21 shows the penalty in MDL estimation for different coupling regimes. Similar to the MD vector estimation, the penalty in MDL estimation decreases as we approach the strong-coupling regime.

Since MD and MDL of the fiber do not add up linearly for different fiber lengths, we compute the norm of the MD vector for different numbers of modes and varying fiber lengths

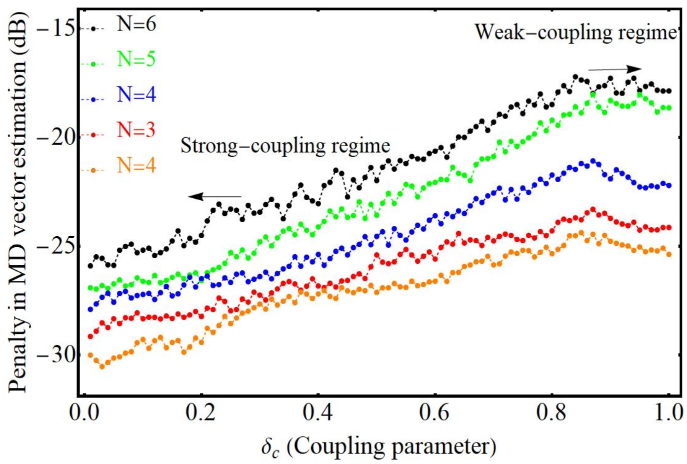

Figure 20: Penalty in the MD vector estimation as a function of coupling parameter (Symbols: Orange: N=2, Red: N=3, Blue: N=4, Green: N=5, Black: N=6).

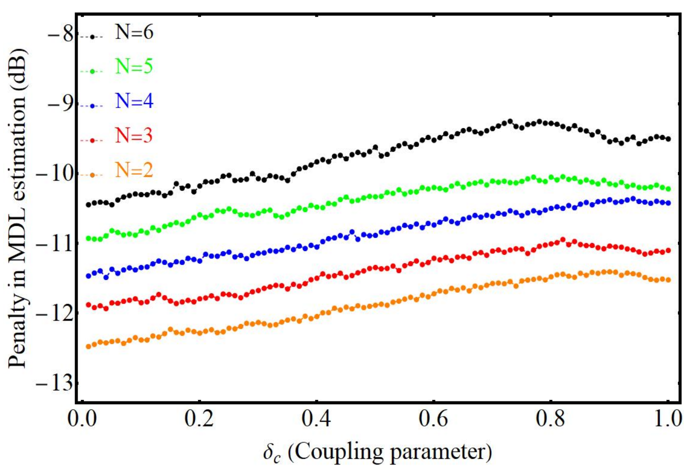

Figure 21: Penalty in the MDL estimation as a function of coupling parameter (Symbols: Orange: N=2, Red: N=3, Blue: N=4, Green: N=5, Black: N=6).

to illustrate the impact of MDL on MD characterization.

Fig. 22 shows the increase in the norm of the MD vector as a function of fiber length for different numbers of modes, in both the absence and presence of MDL. As shown, in the joint presence of MD and MDL, the MDL induces progressively larger penalties in MD vector estimation as the fiber length increases. For instance, for a 20 km long 6-mode fiber, the MD is measured with more than 1.6 ps error compared to the case without MDL.

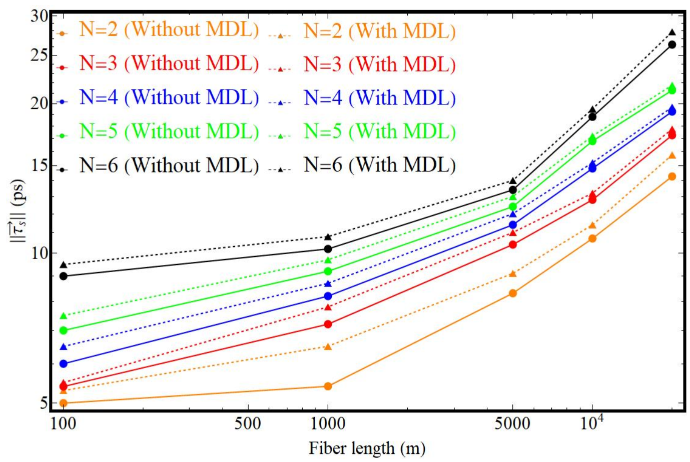

Figure 22: MD vector vs. fiber length for different number of modes in the absence and presence of MDL.

Since the principal attenuation of the fiber determines the MDL and accumulates over the fiber length, the MDL increases as the fiber length increases.

Fig. 23 illustrates the MDL of the fiber as a function of fiber length for different numbers of modes. For example, in a 6-mode fiber, the MDL increases by an average of 0.0625 dB for every 1 km of fiber.

Moreover, MDL can also be introduced by imperfections in transmission system components, such as mode converters and optical amplifiers.

To investigate the accuracy of the method, we consider additional limiting factors such as modal crosstalk and thermal noise, along with MDL, and evaluate their combined impact on channel performance. The penalty in MD vector estimation as a function of SNR for different sources of error is shown in Fig. 24. In this figure, all sources of error—thermal noise at the direct-detection receiver, MDL as a transmission impairment of the fiber, and mode combiner error—are taken into account to assess the accuracy of the MD-SDM and the MD-APM. As shown, MDL introduces a higher penalty in MD vector estimation compared to modal crosstalk. In the presence of thermal noise, the penalties due to MDL and crosstalk accumulate, further degrading estimation performance.

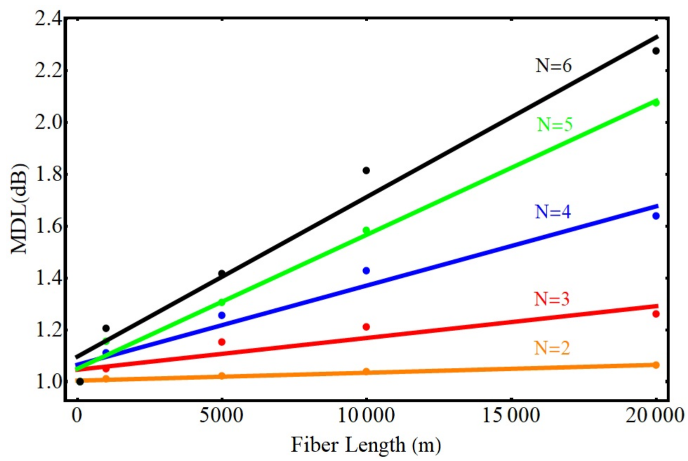

Figure 23: MDL vs. fiber length for different number of modes (Symbols: Orange: =2, Red: =3, Blue: =4, Green: =5, Black: =6).

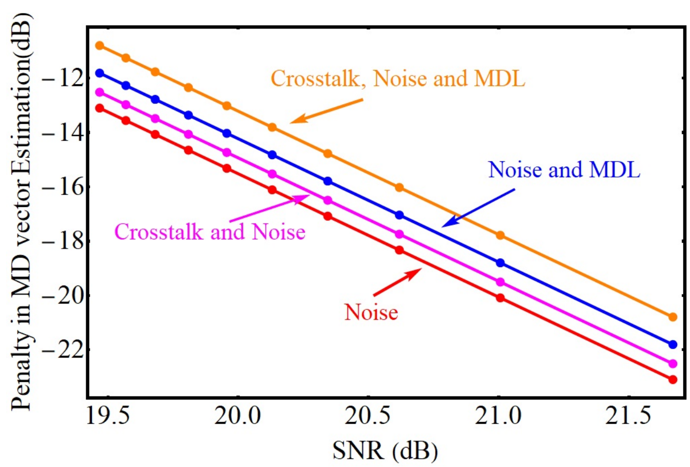

Figure 24: Penalty in MD vector estimate as a function of the SNR to assess different source of errors (Symbols: Red: Impact of noise, Magenta: Impact of crosstalk and noise, Blue: Impact of noise and MDL, Orange: Impact of crosstalk, Noise, and MDL for a 2-mode fiber).

As mentioned before, using optimal launch vectors can minimize the error in MD vector estimation. Fig. 25 compares different input launch vectors, such as Yang and Nolan's vectors [94], mutually unbiased bases (MUBs) [95], symmetric, informationally complete, positive operator-valued measure (SIC-POVM) vectors [96], and the optimal launch vectors proposed in [17].

We observe that the optimal launch vectors from [17] produce a lower penalty compared to other previously proposed vectors. For instance, in a 6-mode fiber, the penalty in MD estimation is reduced by 3.22 dB when using optimal launch vectors instead of random launch vectors, under a modal crosstalk level of -30 dB at the fiber input.

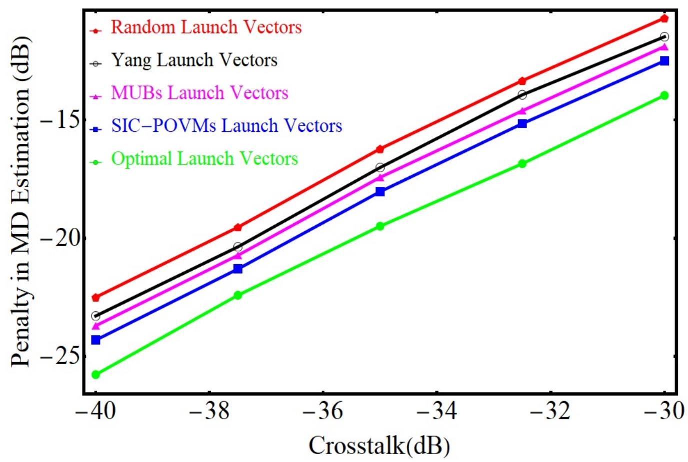

Figure 25: Penalty in MD vector estimate as a function of crosstalk for different launch vectors (Symbols: Red: Random launch vectors, Black: Yang and Nolan's vectors, Magenta: MUBs launch vectors, Blue: SIC-POVMs launch vectors, Green: Optimal launch vectors).

Fig. 26 illustrates the effect of using different launch vectors for MDL measurement as a function of crosstalk. As expected, the optimal launch vectors [17] minimize the error in MDL estimation. In fact, these optimal vectors reduce the penalty by 4.11 dB compared to random launch vectors when the modal crosstalk at the fiber input is -40 dB.

To enhance the accuracy of MD-SDM and MD-APM fiber characterization techniques, this study proposes an optimization approach that maximizes the volume of the parallelotope formed by launch Stokes vectors. Through gradient ascent algorithms—initialized using different vector sets such as Yang and Nolan's vectors, MUBs, and SIC-POVMs—it was shown that vectors with high symmetry, particularly SIC-POVMs, enable faster convergence and better orthogonality in Stokes space. The Gram matrix analysis confirmed that

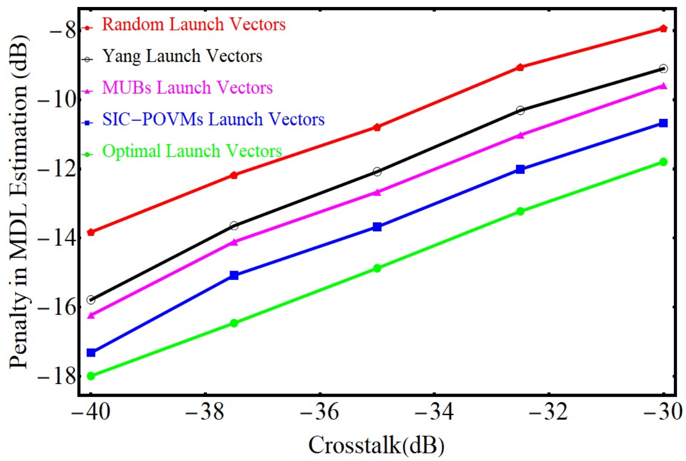

Figure 26: Penalty in MDL estimation as a function of crosstalk for different launch vectors (Symbols: Red: Random launch vectors, Black: Yang and Nolan's vectors, Magenta: MUBs launch vectors, Blue: SIC-POVMs launch vectors, Green: Optimal launch vectors).

the optimization process yields nearly orthogonal vector sets, minimizing obliqueness and improving the estimation accuracy. Additionally, results demonstrated that maximizing the parallelotope volume provides a practical and computationally efficient alternative to noise-based cost function minimization, yielding only a minor performance trade-off in terms of SNR penalty.

Extensive simulations validated the performance of MD-SDM and MD-APM under various transmission conditions. These methods successfully estimated MD and MDL for MMFs and MCFs, though their accuracy was influenced by impairments such as MDL, thermal noise, and modal crosstalk. Mode coupling, fiber length, and launch vector selection were shown to significantly impact estimation penalties. Strong coupling regimes reduced MD and MDL estimation errors, while longer fiber lengths amplified them. Optimal launch vectors, derived through volume maximization, significantly reduced estimation penalties—by up to 4.11 dB in MDL estimation compared to random vectors—especially under high crosstalk. Overall, the results emphasize the importance of robust launch vector design and impairment-aware characterization to enable accurate and scalable deployment of SDM fibers in next-generation optical communication systems.

DSP algorithms performance

In this section, results of an iterative OEKF are provided. We iterated the OEKF for 35 Stokes vectors and attempted to de-rotate the rotated state of polarization (SOP) of each vector for a 6-mode fiber. Then, we stored the Euclidean norm of the distances as an error between the rotated and target SOPs, defined as

$$Error = \|X - Y\|. \tag{5.1}$$

As shown in Fig. 27, the penalty is plotted as a function of the number of iterations for various step sizes. Larger step sizes lead to higher penalties as the algorithm progresses through iterations, whereas smaller step sizes result in a substantial decrease in penalty as the number of iterations increases. We observe that a step size of 0.000087 yields the most gradual and consistent reduction in penalty. This highlights the importance of selecting an appropriate step size to achieve optimal EKF performance in a 6-mode fiber system.

As mentioned earlier, the measurement noise covariance matrix is an integral part of the Kalman filter, as it indicates the uncertainties in the measurements or observations. During the update phase of the state estimation process, this matrix significantly influences the weighting between the measurements and the predicted state.

Fig. 28 illustrates the penalty as a function of the number of iterations required to reach convergence, or a sufficiently small error between SOPs. It is observed that for larger values of $\mathbf{R}$, the initial penalties are higher and the convergence rate is slower. Conversely, smaller values of $\mathbf{R}$ lead to faster convergence and lower final penalties, with the lowest penalty achieved for $\mathbf{R} = 0.000034\mathbf{I}$.

This result shows the importance of optimizing the measurement noise covariance matrix in order to achieve accurate and efficient state estimation using the OEKF.

The process noise covariance is another important parameter within a Kalman filter

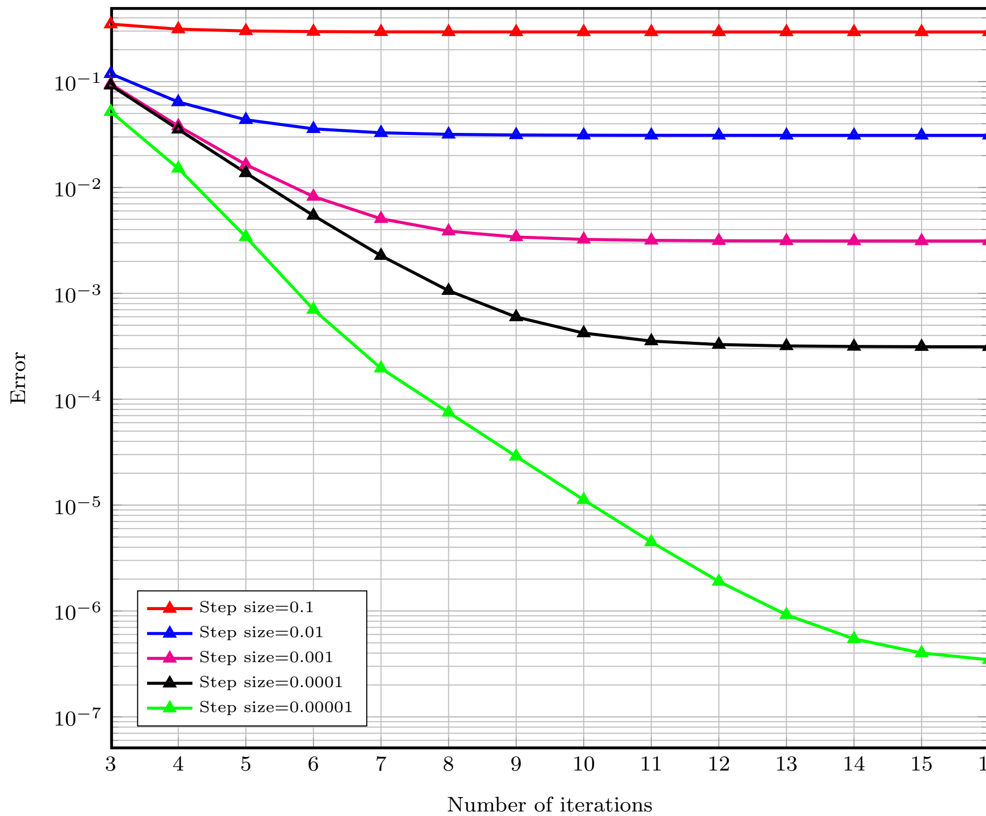

Figure 27: Impact of step size.

that characterizes the uncertainty in the process or system dynamics. It defines the random nature of the change in state variables and how uncertain those changes are over time. The process noise enters during the prediction step of the Kalman filter, where the estimated state is propagated forward based on the system dynamics.

In Fig. 29, the penalty is plotted over the number of iterations for different values of the process noise covariance, $\mathbf{Q}$. From the figure, it can be observed that with higher values of $\mathbf{Q}$, the penalty initially decreases more rapidly. However, this advantage eventually diminishes, as higher $\mathbf{Q}$ values lead to less stable convergence compared to moderate values such as $\mathbf{Q} = 0.1\mathbf{I}$. On the other hand, with smaller values of $\mathbf{Q}$, the decrease in penalty is more gradual, but the lowest final penalties are achieved with very small $\mathbf{Q}$.

This indicates that selecting an optimal process noise covariance is essential for

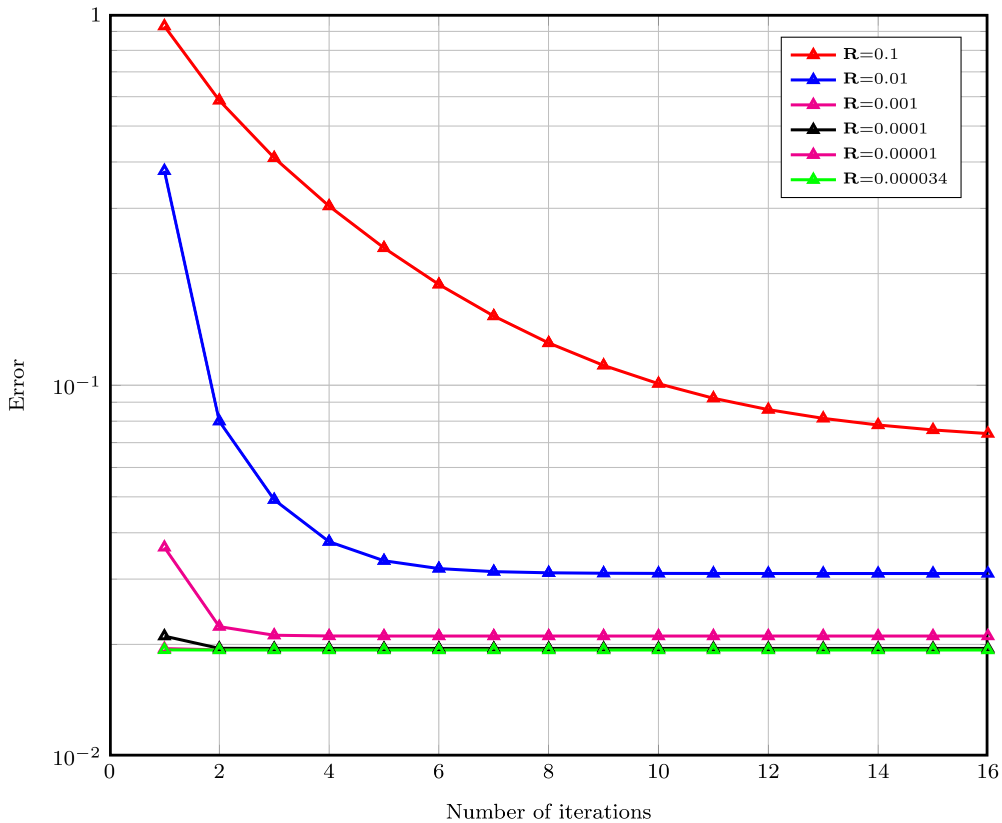

Figure 28: Impact of measurement noise covariance matrix.

achieving the best performance in the state estimation process of a Kalman filter.

It incorporates the state estimation uncertainty by modifying the error covariance matrix and modeling the process noise as a zero-mean random variable with a corresponding covariance matrix.

Fig. 30 shows how a higher initial error covariance causes the penalty to drop rapidly at first, then more gradually until it stabilizes. In contrast, smaller initial error covariances result in a more consistent decrease in penalty throughout the iteration process. Among the different cases, $\mathbf{P} = 0.068\mathbf{I}$ yields both the smallest initial error covariance and the lowest final penalty values.

In our investigation into the performance of electronic equalization techniques for

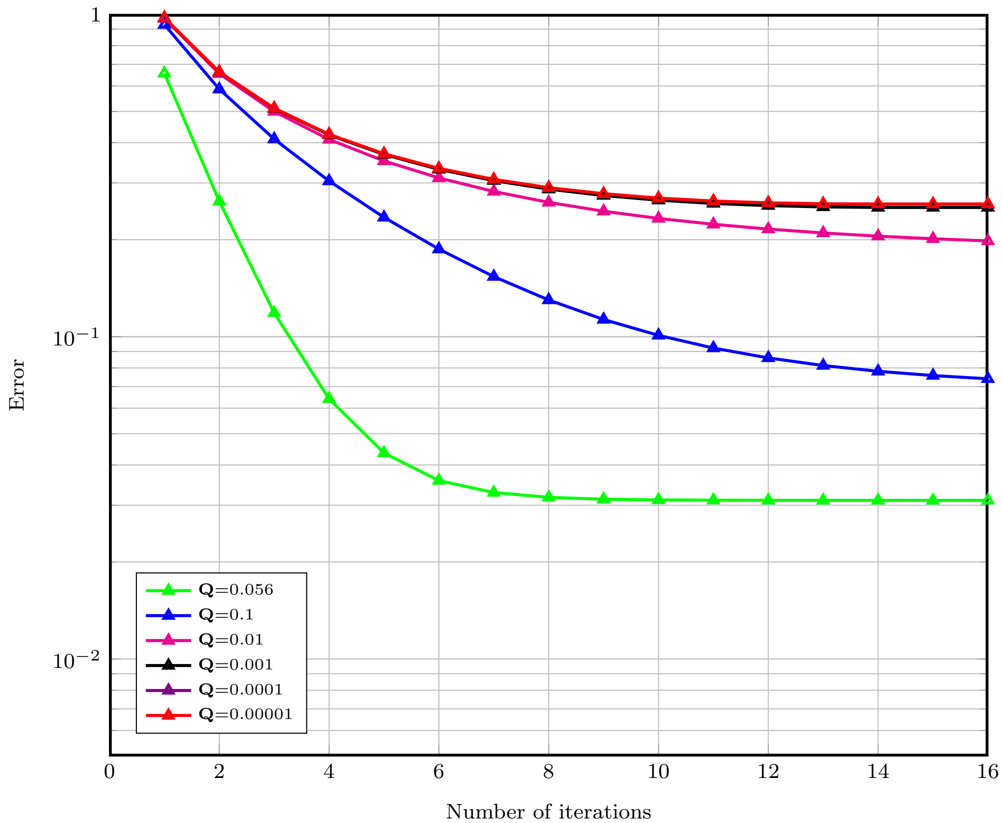

Figure 29: Impact of process noise covariance.

MVM/DD, we conducted comprehensive numerical simulations of BER versus SNR across various scenarios involving different fiber lengths, baud rates, and constellation sizes. Our primary objective was to assess the effectiveness of different equalization methods in mitigating transmission impairments.

We used SVM to visualize the effects of different DSP algorithms on the transmitted signals. Fig. 31 shows the locations of 8 constellation points on the Poincaré sphere before and after transmission through an optical fiber. The blue points indicate the theoretical target positions of the signals after transmission and serve as a reference for evaluating the accuracy and performance of each equalization method.

Fig. 31(a) shows the constellation points on the Poincaré sphere before transmission.

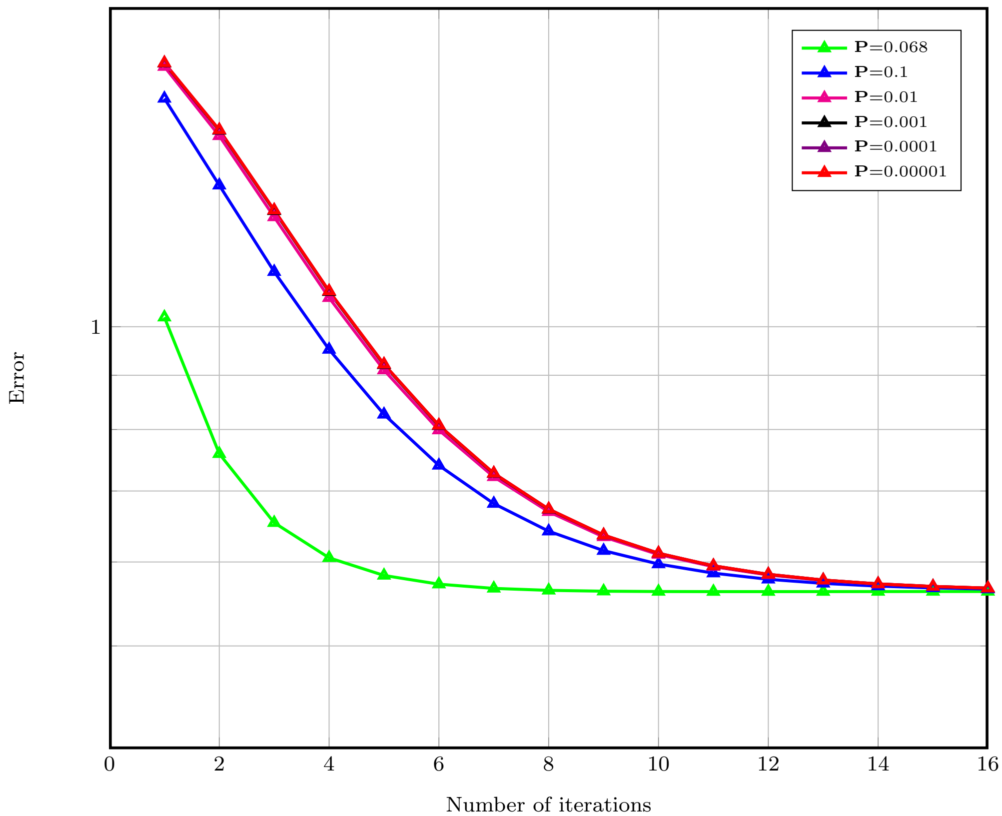

Figure 30: Impact of error covariance.

In this figure, each blue point represents an ideal MVM symbol position, which serves as the target location that the received signal should ideally match after transmission. However, due to intrinsic distortions introduced by the optical fiber and amplified spontaneous emission (ASE) noise from optical pre-amplifiers, the actual received signals deviate from these ideal positions.

As shown in Fig. 31(b), the red points represent the signals received after transmission through the fiber. Clearly, the signals are significantly spread around the ideal blue points, indicating the presence of severe transmission impairments. This visual evidence underscores the challenge of maintaining signal integrity in the presence of fiber-induced distortions and noise.

Fig. 31(c) shows the received signals after applying the LMS equalizer. The symbols

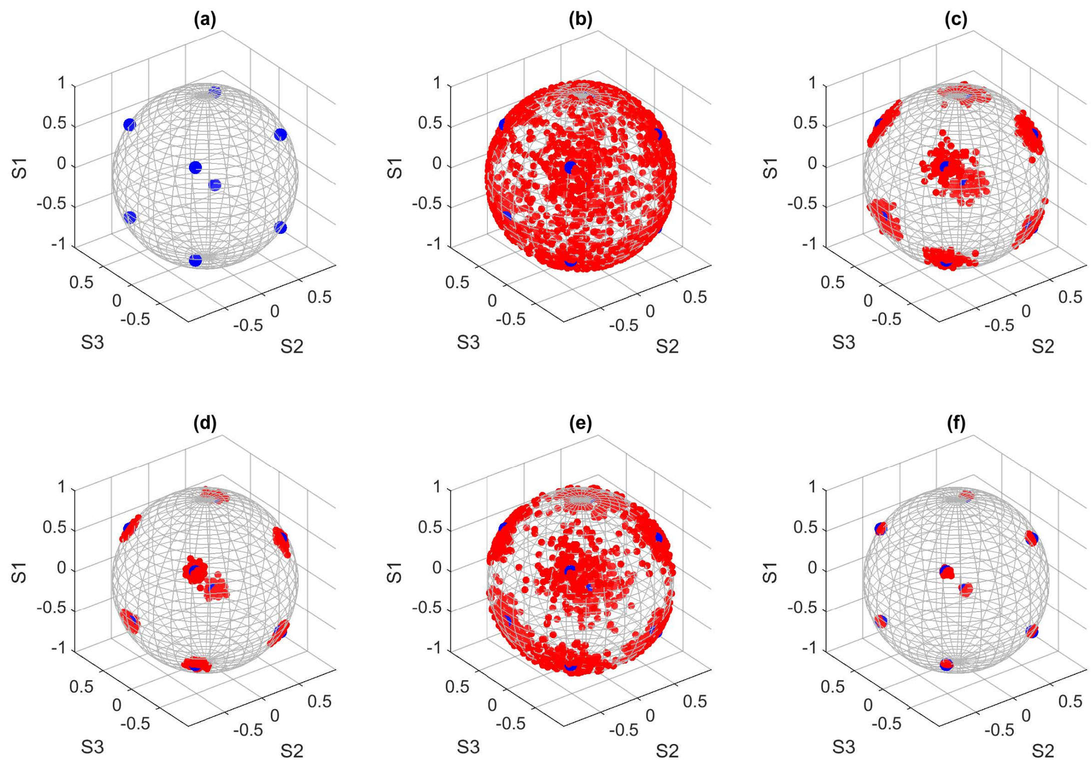

Figure 31: Constellation points for 8-SVM (a) before transmission; (b) after transmission; (c) after LMS; (d) after EKF; (e) after ICA; (f) after OEKF.

are now closer to the target points, but there is still considerable spread, indicating that the LMS equalizer partially compensates for the RSOP, yet it is not sufficient for optimal signal recovery.

Application of the EKF, as shown in Fig. 31(d), results in better convergence toward the blue points compared to the LMS equalizer, demonstrating that the EKF is more effective and accurate in correcting signal distortions.

Fig. 31(e) illustrates the effect of the ICA equalizer on the received signals. It is evident that the ICA equalizer is less effective than both the LMS and the EKF in mitigating distortion.

Finally, Fig. 31(f) presents the performance of the EKF equalizer after optimization. The red points nearly overlap with the blue target points, showing minimal residual spread in the symbol positions on the Poincaré sphere. By effectively mitigating transmission-induced distortions, the OEKF equalizer outperforms all other tested algorithms, resulting in near-ideal signal recovery for an 8-SVM signal.

We simulated several DSP algorithms across various fiber lengths and symbol rates, and the results are depicted in Figures 32–35. Fig. 32 shows the bit error rate (BER) versus signal-to-noise ratio (SNR) for a 0.5 km 6-mode FMF with a differential mode delay (DMD) of 2.5 ps/km, operating at a symbol rate of 130 GBd.

We plot the performance of the DSP algorithms studied—including the LMS equalizer, EKF, ICA equalizer, and OEKF—and compare them with the back-to-back (B2B) configuration. The horizontal red line indicates the forward error correction (FEC) threshold. The B2B configuration serves as a baseline, achieving very low BER as SNR increases.

The LMS equalizer exhibits a higher BER that gradually improves with increasing SNR, though it remains above the FEC threshold. The EKF outperforms the LMS equalizer, reaching below the FEC threshold only at higher SNR levels. The ICA equalizer delivers intermediate performance, staying above the FEC threshold for lower SNR values. While EKF performs better than LMS and ICA, the best performance is achieved by the OEKF, demonstrating its effectiveness in equalizing signals by de-rotating the received polarization states.

We plot the performance of different equalizers for a 2.5 km fiber with a symbol rate of 57 GBd in Fig. 33. It can be observed that the OEKF provides the best performance, achieving the lowest BER across the SNR range. The LMS and ICA algorithms yield relatively poorer performance compared to the Kalman filter-based algorithms. While the EKF performs better than LMS and ICA, it still falls short of the OEKF, particularly under conditions of severe DMD.

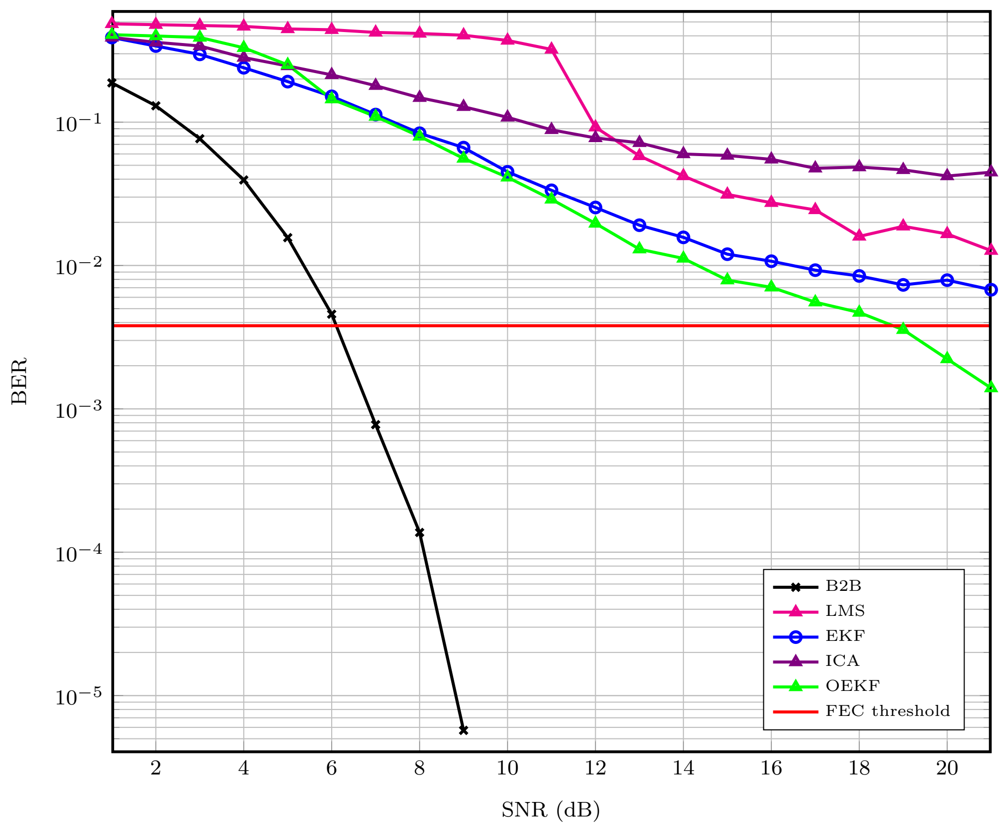

Figure 32: BER vs. SNR (dB), L=0.5 km, DMD=2.5 ps/km, Symbol rate=130 GBd, (6,64)-MVM.

In Fig. 34, we compare the performance of different equalizers for a 10 km fiber operating at a symbol rate of 29 GBd. The OEKF reduces the BER below the FEC threshold at an SNR of 17 dB, demonstrating its effectiveness in de-rotating the MVM signals and mitigating transmission impairments.

As shown in Fig. 35, we increased the fiber length to 40 km with a symbol rate of 14 GBd and evaluated the performance of the equalizers. The results indicate that the OEKF achieves the best performance among all equalizers, demonstrating its robustness even in long-haul fiber scenarios.

In Figures 36–39, we explored the performance of the algorithms as a function of various constellation cardinalities. The performances for cardinality $M = 32$ and $M = 64$ are shown

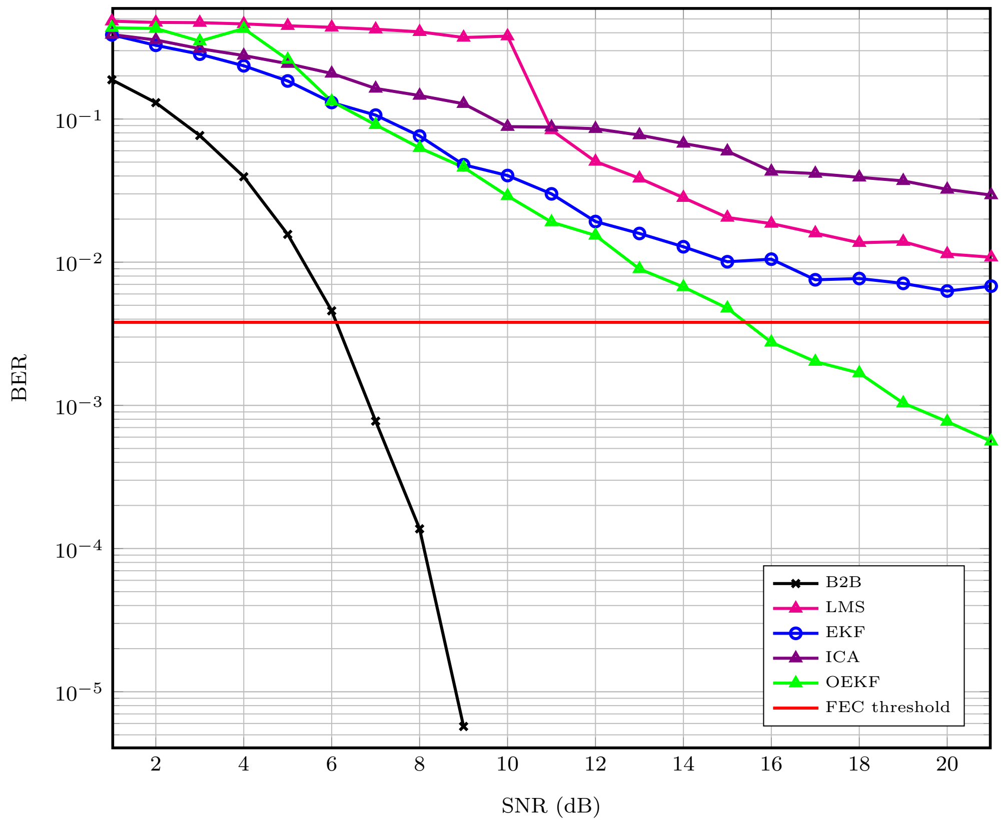

Figure 33: BER vs. SNR (dB), L=2.5 km, DMD=2.5 ps/km, Symbol rate=57 GBd, (6,64)-MVM.

in Figures 36 and 37, respectively. In both cases, the OEKF efficiently equalizes and de-rotates the signal, achieving BER levels that meet the FEC threshold.

However, as the constellation cardinality further increases, as shown in Figures 38 and 39, the OEKF also fails to fully equalize and de-rotate the signals. These figures clearly demonstrate the limitations of the optimized Kalman filter at higher cardinalities, as the BER remains above the FEC threshold even with increased SNR.

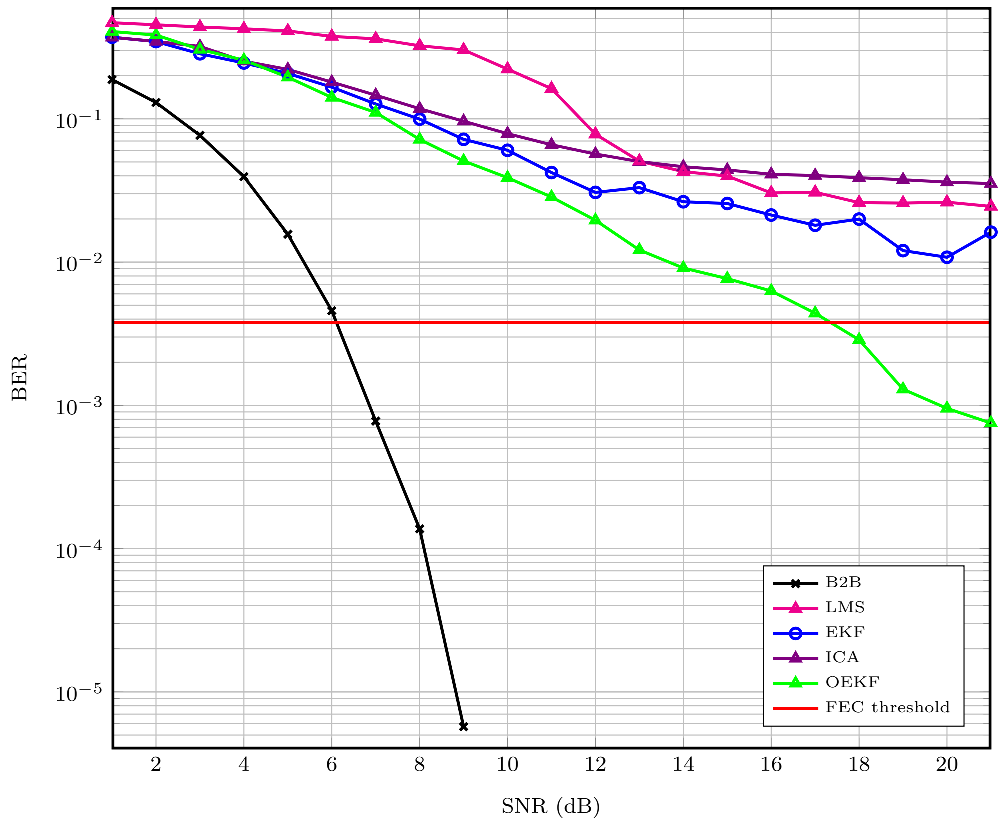

Figure 34: BER vs. SNR (dB), L=10 km, DMD=2.5 ps/km, Symbol rate=29 GBd, (6,64)-MVM.

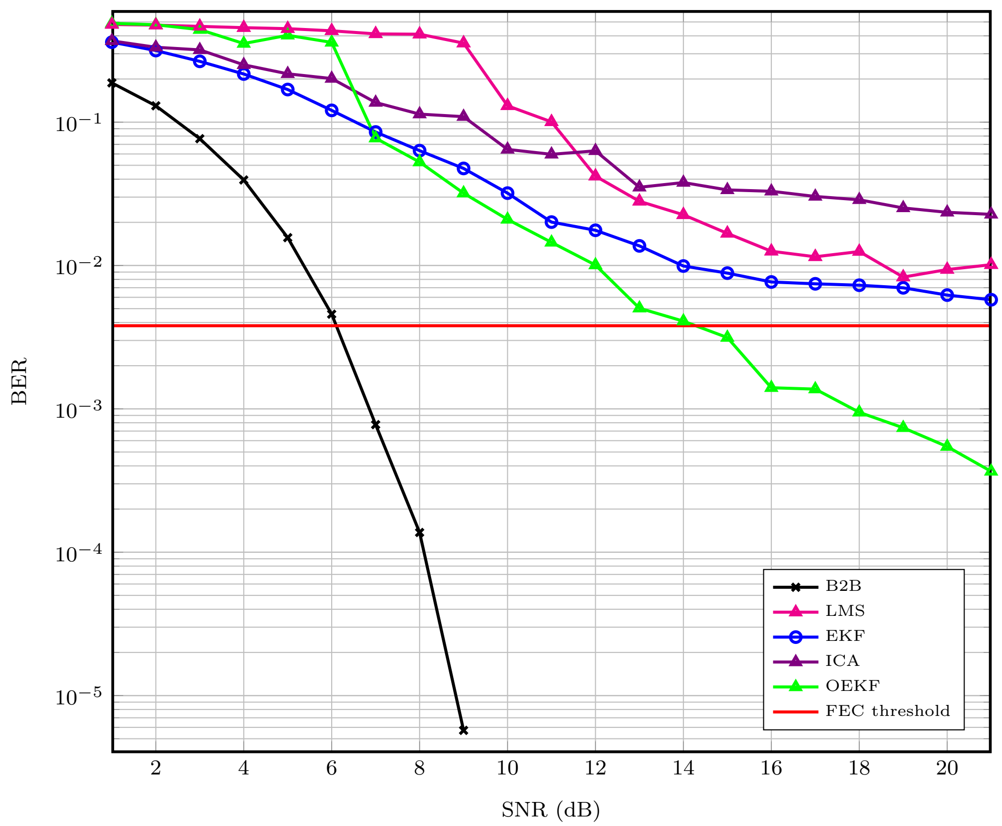

Figure 35: BER vs. SNR (dB), L=40 km, DMD=2.5 ps/km, Symbol rate=14 GBd, (6,64)-MVM.

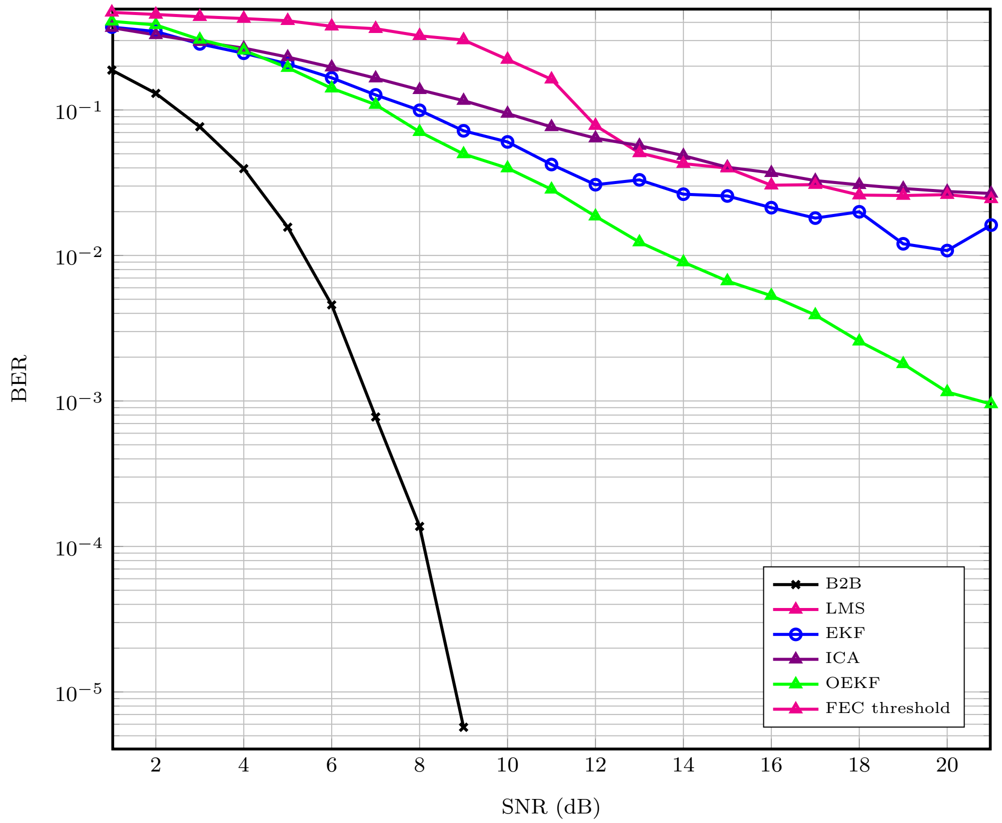

Figure 36: BER vs. SNR (dB) for (6,32)-MVM.

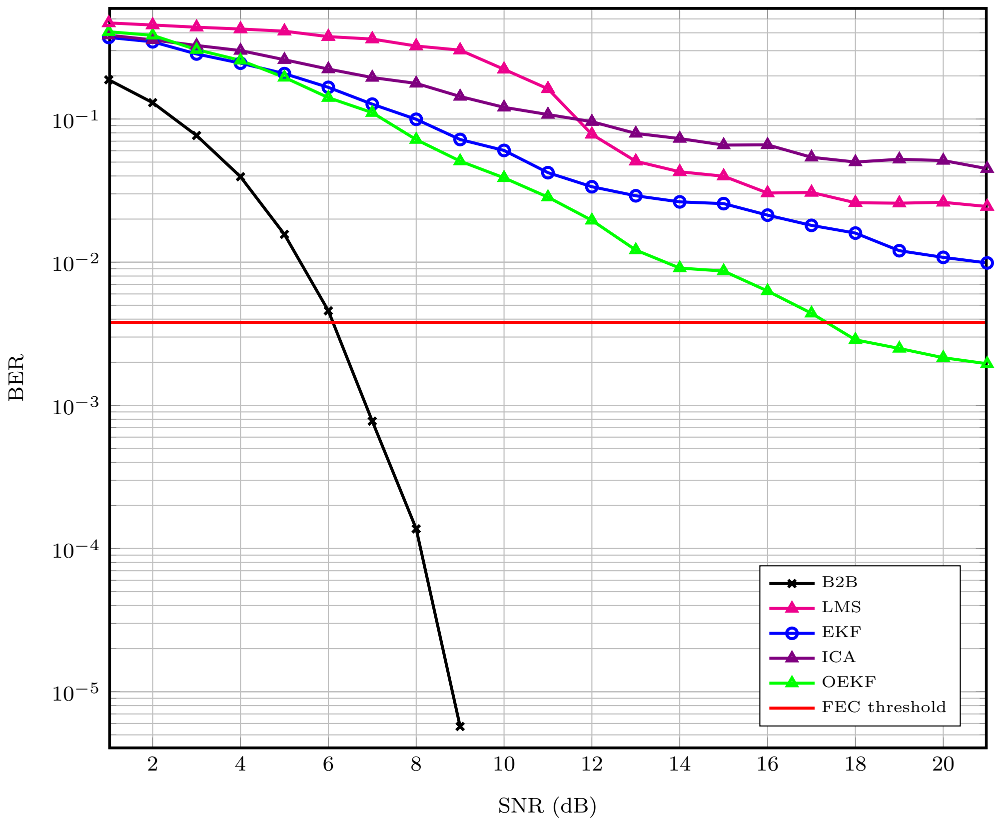

Figure 37: BER vs. SNR (dB) for (6,64)-MVM.

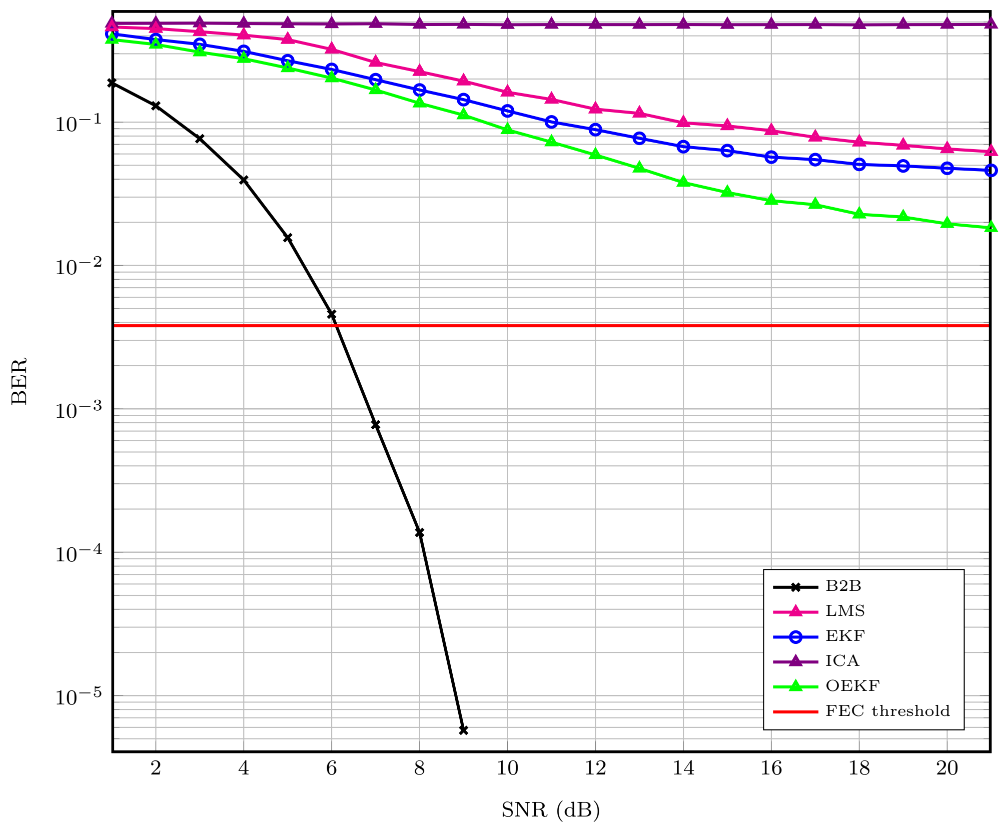

Figure 38: BER vs. SNR (dB) for (6,256)-MVM.

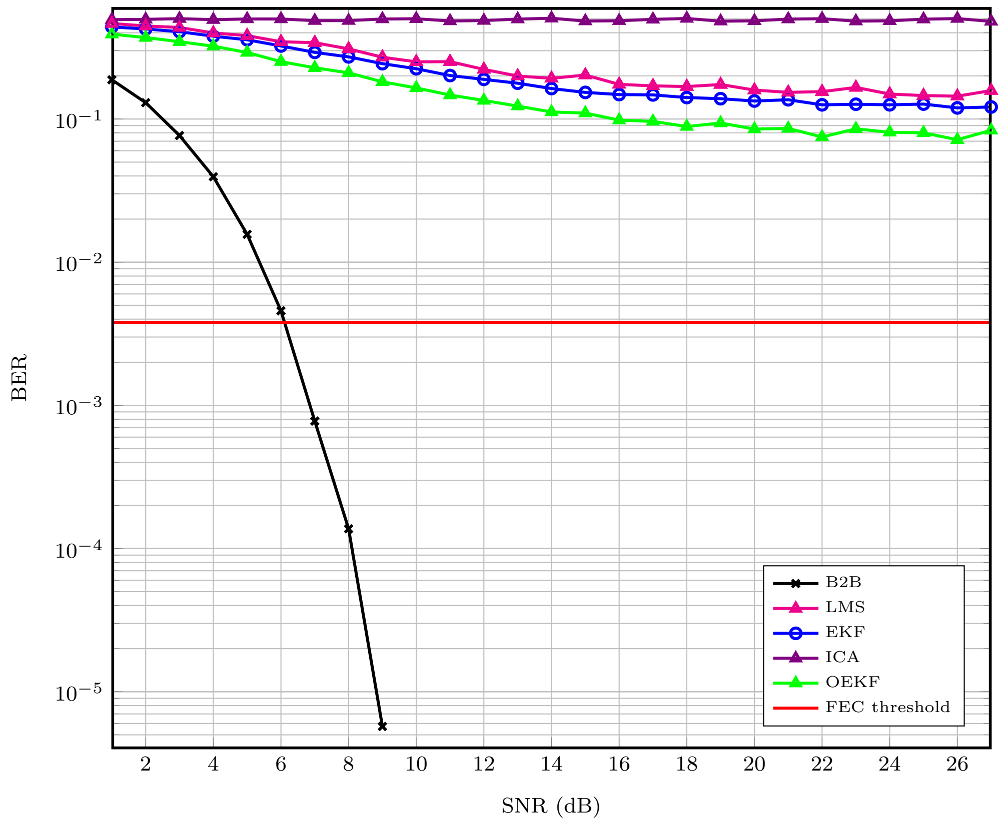

Figure 39: BER vs. SNR (dB) for (6,1024)-MVM.

Fig. 40 compares the performance of the OEKF for different constellation cardinalities with the B2B case, which serves as a baseline and consistently exhibits the lowest BER for all SNR values, representing ideal performance in the absence of transmission impairments.

The OEKF shows varying levels of effectiveness in BER reduction as the constellation cardinality increases. For $M = 32$ and $M = 64$, the OEKF achieves BER values below the FEC threshold at higher SNR levels, indicating successful equalization and de-rotation of the signal. However, for higher cardinalities, all equalizers—including the OEKF—fail to compensate for the RSOP of the fiber, and the BER remains above the FEC threshold.

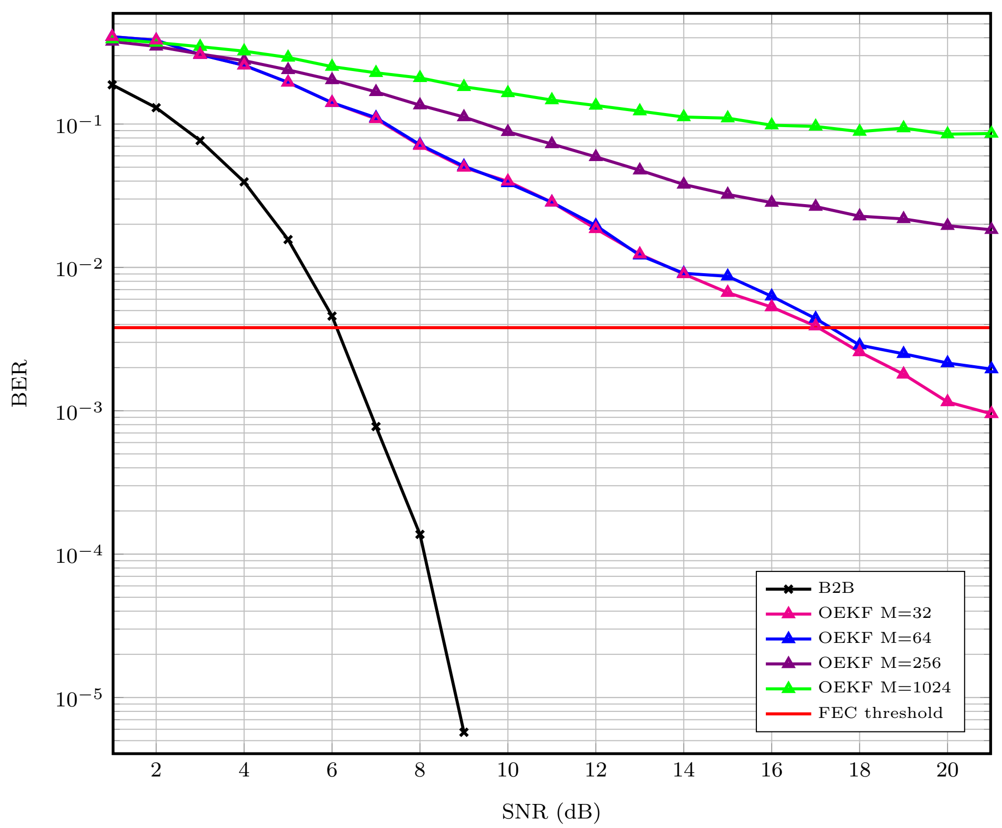

Figure 40: BER vs. SNR (dB) for various cardinality values.

Fig. 41 shows the BER versus SNR for different levels of crosstalk, ranging from –50 dB/km to –10 dB/km. The results indicate that varying the level of crosstalk has minimal impact on the rotation of the state of polarization. As a result, all equalizers are able to efficiently equalize the signal across different crosstalk levels in the absence of other impairments.

This suggests that, in scenarios without chromatic dispersion (CD) and differential mode delay (DMD), the equalizers demonstrate strong robustness in mitigating the effects of crosstalk.

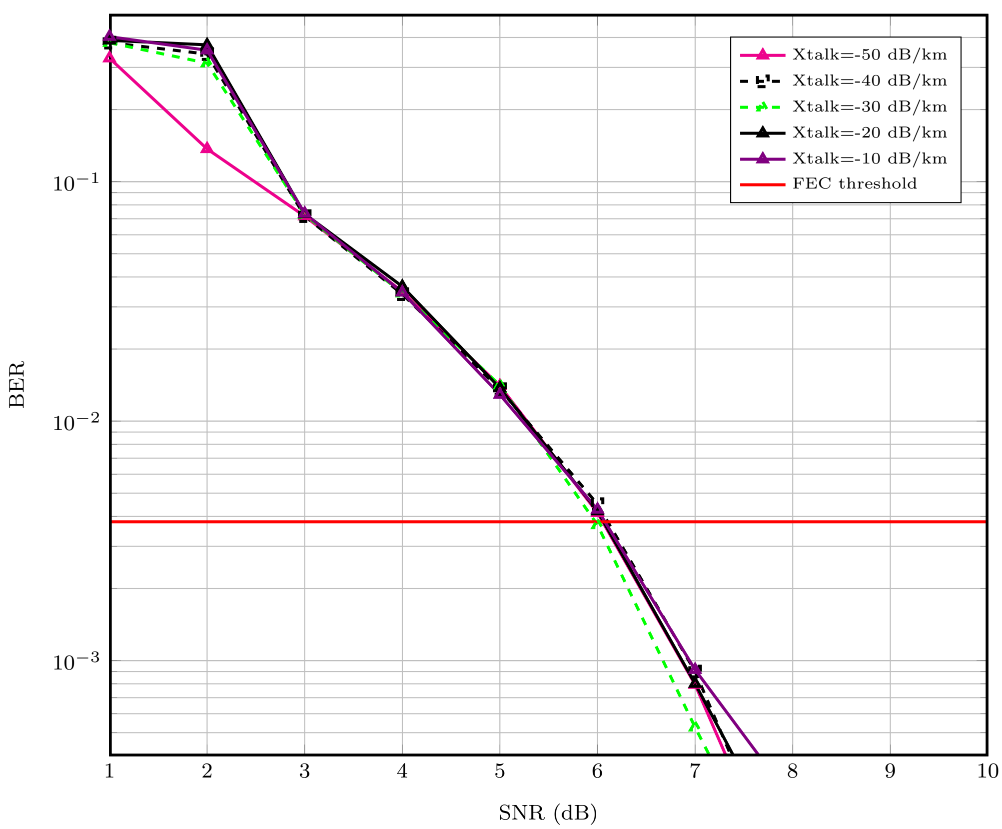

Figure 41: BER vs. SNR (dB) for various crosstalk levels.

In Fig. 42, we compare the performance of different symbol rates: 28, 56, and 112 GBd. It is observed that the SNR penalty increases with higher symbol rates, necessitating more advanced DSP techniques. When using the OEKF, the SNR penalty is reduced by 2.5 dB at 28 GBd and 5.4 dB at 112 GBd compared to the ICA method. Additionally, the OEKF shows a reduction of 0.7 dB and 3.3 dB compared to the LMS method at the same symbol rates, respectively.

This comparison is performed under low DMD conditions, where all equalizers function effectively. However, under higher levels of DMD, only the OEKF equalizer is capable of successfully equalizing the received signal. As a result, direct comparisons of SNR penalty with other equalizers are not feasible under those conditions.

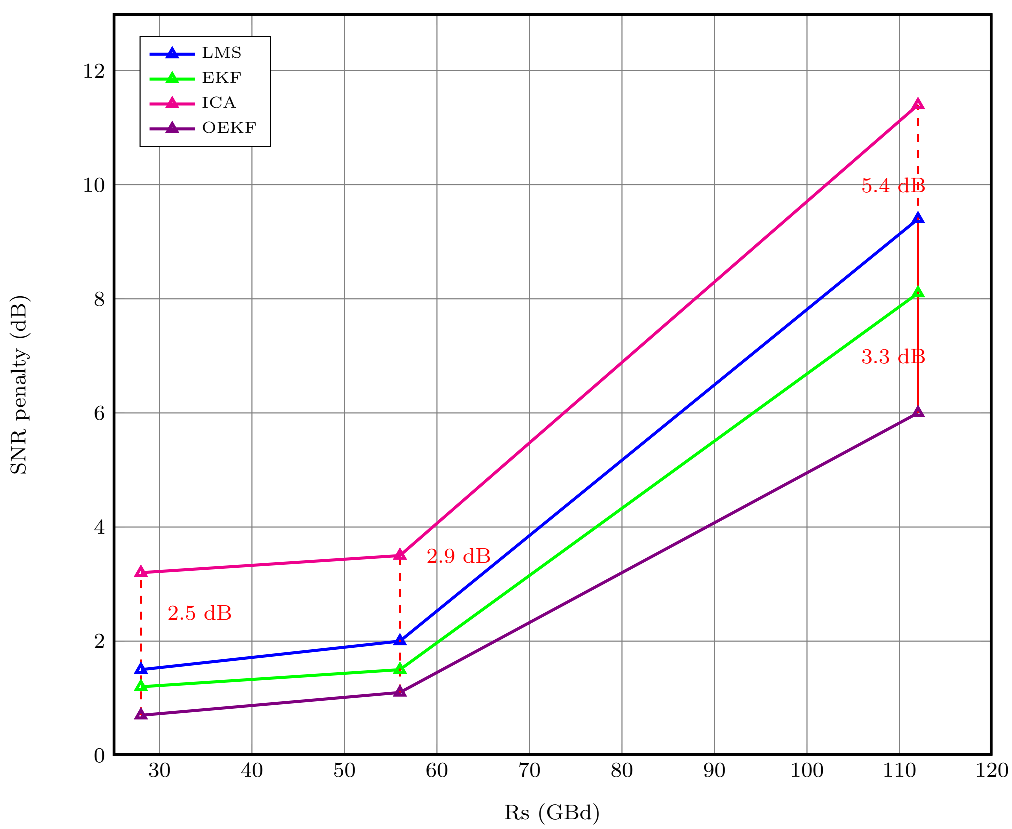

Figure 42: SNR penalty vs. Rs.

Fig. 43 shows the percentage of fiber reach extension as a function of symbol rate when using the EKF and OEKF compared to the LMS method. The results indicate that, at a symbol rate of 112 GBd, the EKF and OEKF extend the fiber reach by 31% and 49%, respectively. These findings demonstrate that the OEKF significantly outperforms the other equalizers at higher symbol rates, making it particularly effective for high-speed optical communication systems.

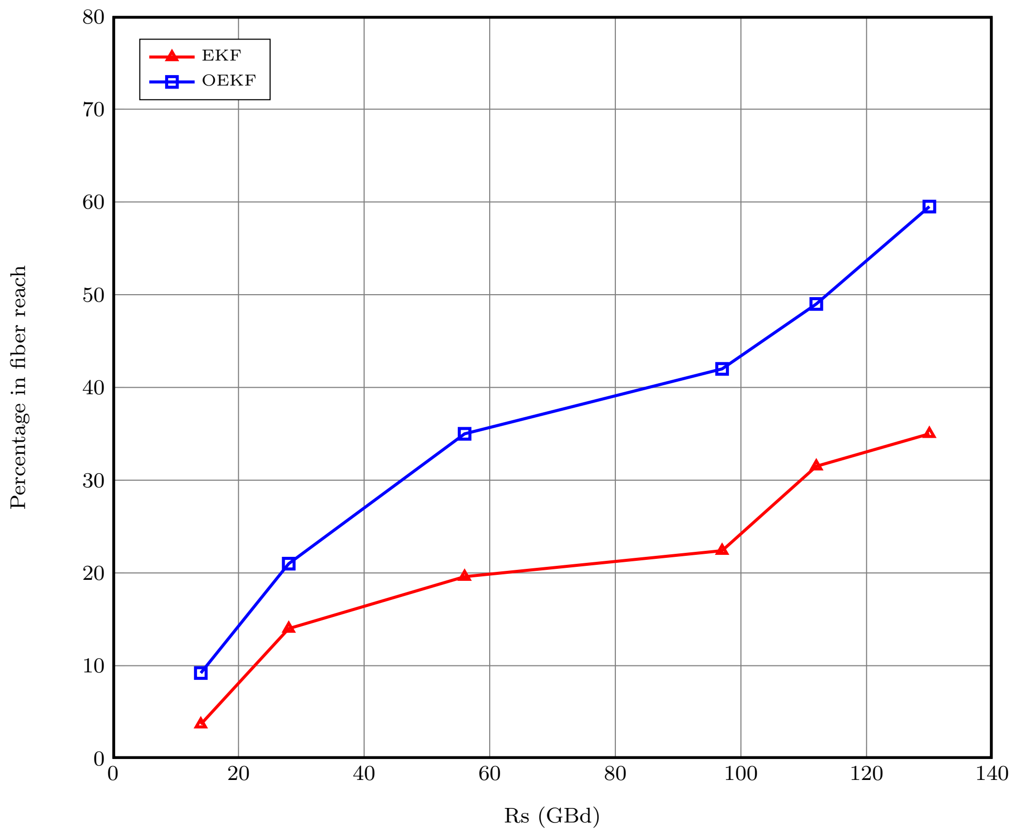

Figure 43: Percentage in fiber reach vs. Rs.

Overall, The LMS and ICA equalizers offer relatively simple and computationally efficient solutions for equalizing MVM/DD signals but fall short in scenarios involving high symbol rates, longer fiber lengths, or severe polarization rotation. LMS relies on a gradient descent-based adaptive algorithm, making it suitable for moderate distortion and low-complexity systems. However, its convergence is slow and less accurate in highly dynamic environments. ICA, while capable of blind source separation, is less effective in tracking fast-varying polarization states and suffers from higher BER, especially under strong transmission impairments. Simulation results show that both LMS and ICA fail to bring the BER below the FEC threshold in high-data-rate or high-order modulation scenarios, highlighting their limitations for advanced optical systems.

In contrast, the EKF and OEKF are more robust and accurate due to their recursive state estimation capabilities, which dynamically adapt to the channel's nonlinearities and time-varying characteristics. EKF outperforms both LMS and ICA by effectively reducing the BER under moderate conditions, though its performance starts to degrade at higher constellation sizes or symbol rates. The OEKF further enhances the standard EKF by incorporating adaptive optimization of filter parameters, allowing for faster convergence, lower BER, and greater resilience to impairments such as DMD and RSOP. It consistently outperforms all other equalizers across a wide range of scenarios—including long-haul transmission and high-order modulation—extending fiber reach by up to 49% at 112 GBd and achieving near-ideal constellation recovery on the Poincaré sphere. This makes the OEKF the most capable equalizer among those studied, particularly in high-performance optical communication systems.

CONCLUSION

The findings of this dissertation contribute to the advancement of SDM-based optical communication systems by addressing key challenges related to fiber characterization, modulation formats, and DSP algorithms at the receiver. With the exponential growth in global data traffic, innovative solutions are necessary to maximize the capacity of optical fibers. This study introduces approaches that support the effective deployment of SDM technology in high-capacity communication networks through improved fiber characterization techniques, enhanced modulation efficiency, and optimized DSP algorithms for impairment compensation.

The MD-SDM and MD-APM methods were investigated for characterizing MMFs and MCFs, focusing on the accurate estimation of MD and MDL—two critical parameters that significantly influence SDM system performance. Monte Carlo simulations revealed that the accuracy of these methods is sensitive to transmission impairments such as thermal noise, modal crosstalk, and mode coupling. To improve estimation accuracy, optimal launch vectors were computed by maximizing the volume of the parallelotope formed by the Stokes vectors. By comprehensively evaluating SDM fiber characterization under varying transmission conditions, this research highlights the importance of robust and precise characterization techniques to ensure the practical viability of SDM fibers for high-speed, high-capacity optical communication applications.

In addition to fiber characterization, this study investigates multi-dimensional modulation formats designed to optimize data transmission in SDM systems. Mode Vector Modulation (MVM) was analyzed under various transmission impairments, including rotation of the state of polarization (RSOP), differential mode delay (DMD), chromatic dispersion (CD), and modal crosstalk. Monte Carlo simulations revealed that MVM offers improved spectral efficiency and supports higher data transmission rates compared to conventional modulation formats. Furthermore, the integration of SDM with wavelength-division multiplexing (WDM) was explored, demonstrating that a combined SDM-WDM architecture can significantly enhance overall system performance by exploiting both spatial and spectral dimensions. These findings underscore the viability of MVM as a scalable and efficient modulation format for next-generation optical networks.

A major contribution of this research lies in the development and evaluation of digital signal processing (DSP) algorithms for mitigating transmission impairments in SDM systems. Several DSP techniques—namely LMS, ICA, the Kalman filter, and the Extended Kalman Filter (EKF)—were implemented and systematically compared based on their ability to compensate impairments in MVM-based direct detection transceivers. Among these, the proposed Optimized EKF (OEKF) demonstrated superior performance in terms of convergence, bit error rate (BER), and robustness under various impairment scenarios. By integrating these advanced equalization strategies, this work contributes to the development of adaptive, high-performance DSP solutions that are essential for the reliable deployment of high-capacity SDM-based optical communication systems.

In general, this research aims to improve SDM system performance through enhanced fiber characterization and advanced DSP techniques. The methodologies and findings presented here contribute to the practical deployment of SDM technology by addressing critical challenges that must be overcome before its widespread commercial adoption. By ensuring accurate fiber characterization and developing efficient DSP algorithms, this work paves the way for high-capacity optical networks capable of meeting the growing demands of modern communication systems.

Future work could focus on the development of advanced machine learning-based DSP techniques, the exploration of novel multi-dimensional modulation schemes, and the investigation of hybrid integration approaches to further optimize the efficiency, scalability, and robustness of SDM-enabled networks.

REFERENCES

[1] Nokia, "Global Network traffic report, Understanding the growing impact of advancing technologies on future networks," 2023. [Online]. Available: https://onestore.nokia.com/asset/213660.

[2] J. Benjamin *et al.*, "S, C and extended L-band transmission with doped fiber and distributed Raman amplification," in *Proc. IEEE Opt. Fiber Commun. Conf*, 2021.

[3] D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," *Nature Photonics*, vol. 7, no. 5, pp. 354–362, 2013.

[4] P. Winzer, R. Ryf, S. Randel, I. Kaminow, T. Li, and A. Willner, "Spatial multiplexing using multiple-input multiple-output signal processing," *Optical Fiber Telecommunications VIB*, 2013.

[5] P. J. Winzer and D. T. Neilson, "From scaling disparities to integrated parallelism: A decathlon for a decade," *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1099–1115, 2017.

[6] D. Soma, Y. Wakayama, S. Beppu, S. Sumita, T. Tsuritani, T. Hayashi, T. Nagashima, M. Suzuki, M. Yoshida, K. Kasai, M. Nakazawa, H. Takahashi, K. Igarashi, I. Morita, and M. Suzuki, "10.16-Peta-B/s Dense SDM/WDM Transmission Over 6-Mode 19-Core Fiber Across the C+L Band," *Journal of Lightwave Technology*, vol. 36, no. 6, pp. 1362–1368, 2018.

[7] G. Rademacher, B. J. Puttnam, R. S. Luís, J. Sakaguchi, W. Klaus, T. A. Eriksson, Y. Awaji, T. Hayashi, T. Nagashima, T. Nakanishi, T. Taru, T. Takahata, T. Kobayashi, H. Furukawa, and N. Wada, "10.66 peta-bit/s transmission over a 38-core-three-mode fiber," in *Optical Fiber Communication Conference (OFC) 2020*, p. Th3H.1, Optica Publishing Group, 2020.

[8] L. Galdino, A. Edwards, W. Yi, E. Sillekens, Y. Wakayama, T. Gerard, W. S. Pelouch, S. Barnes, T. Tsuritani, R. I. Killey, D. Lavery, and P. Bayvel, "Optical Fibre Capacity Optimisation via Continuous Bandwidth Amplification and Geometric Shaping," *IEEE Photonics Technology Letters*, vol. 32, no. 17, pp. 1021–1024, 2020.

[9] K.-P. Ho, J. M. Kahn, I. Kaminow, T. Li, and A. Willner, "Mode coupling and its impact on spatially multiplexed systems," *Optical Fiber Telecommunications VI*, vol. 17, pp. 1386–1392, 2013.

[10] C. Antonelli, A. Mecozzi, M. Shtaif, and P. J. Winzer, "Stokes-space analysis of modal dispersion in fibers with multiple mode transmission," *Optics Express*, vol. 20, no. 11, pp. 11718–11733, 2012.

[11] I. Roudas and J. Kwapisz, "Stokes space representation of modal dispersion," *IEEE Photonics Journal*, vol. 9, no. 5, pp. 1–15, 2017.

[12] T.-J. Ahn, S. Moon, S. Kim, K. Oh, D. Y. Kim, J. Kobelke, K. Schuster, and J. Kirchhof, "Optical frequency-domain modal dispersion measurement in multimode fibers using intermodal interferometer," in *Optical Fiber Communication Conference*, p. OWI17, Optica Publishing Group, 2006.

[13] J. Nicholson, A. D. Yablon, S. Ramachandran, and S. Ghalmi, "Spatially and spectrally resolved imaging of modal content in large-mode-area fibers," *Optics Express*, vol. 16, no. 10, pp. 7233–7243, 2008.

[14] J. Demas and S. Ramachandran, "Sub-second mode measurement of fibers using C2 imaging," *Optics Express*, vol. 22, no. 19, pp. 23043–23056, 2014.

[15] J. Carpenter, B. J. Eggleton, and J. Schröder, "Observation of Eisenbud–Wigner–Smith states as principal modes in multimode fibre," *Nature Photonics*, vol. 9, no. 11, pp. 751–757, 2015.

[16] G. Milione, D. A. Nolan, and R. R. Alfano, "Determining principal modes in a multimode optical fiber using the mode dependent signal delay method," *Journal of the Optical Society of America B*, vol. 32, no. 1, pp. 143–149, 2014.

[17] I. Roudas, J. Kwapisz, and D. A. Nolan, "Optimal launch states for the measurement of principal modes in optical fibers," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4915–4931, 2018.

[18] M. Dadras, I. Roudas, and J. Kwapisz, "MD and MDL Characterization Using Direct-Detection Pulse Measurements," in *2020 IEEE Photonics Conference (IPC)*, pp. 1–2, IEEE, 2020.

[19] R. Ryf, N. K. Fontaine, M. A. Mestre, S. Randel, X. Palou, C. Bolle, A. H. Gnauck, S. Chandrasekhar, X. Liu, B. Guan, *et al.*, "12× 12 MIMO transmission over 130-km few-mode fiber," in *Frontiers in Optics*, pp. FW6C–4, Optica Publishing Group, 2012.

[20] M. Blau, I. Weiss, J. Gerufi, D. Sinefeld, M. Bin-Nun, R. Lingle, L. Grüner-Nielsen, and D. M. Marom, "Variable optical attenuator and dynamic mode group equalizer for few mode fibers," *Optics Express*, vol. 22, no. 25, pp. 30520–30527, 2014.

[21] N. K. Fontaine, C. R. Doerr, M. A. Mestre, R. R. Ryf, P. J. Winzer, L. L. Buhl, Y. Sun, X. Jiang, and R. Lingle, "Space-division multiplexing and all-optical MIMO demultiplexing using a photonic integrated circuit," in *OFC/NFOEC*, pp. 1–3, IEEE, 2012.

[22] B. Huang, N. K. Fontaine, R. Ryf, B. Guan, S. G. Leon-Saval, R. Shubochkin, Y. Sun, R. Lingle Jr, and G. Li, "All-fiber mode-group-selective photonic lantern using graded-index multimode fibers," *Optics Express*, vol. 23, no. 1, pp. 224–234, 2015.

[23] J. Carpenter and T. D. Wilkinson, "Characterization of multimode fiber by selective mode excitation," *Journal of Lightwave Technology*, vol. 30, no. 10, pp. 1386–1392, 2012.

[24] M. Salsi, C. Koebele, D. Sperti, P. Tran, P. Brindel, H. Mardoyan, S. Bigo, A. Boutin, F. Verluise, P. Sillard, *et al.*, "Transmission at 2× 100Gb/s, over two modes of 40km-long prototype few-mode fiber, using LCOS-based mode multiplexer and demultiplexer," in *Optical Fiber Communication Conference*, p. PDPB9, Optica Publishing Group, 2011.

[25] A. M. Velázquez-Benítez, J. E. Antonio-López, J. C. Alvarado-Zacarías, N. K. Fontaine, R. Ryf, H. Chen, J. Hernández-Cordero, P. Sillard, C. Okonkwo, S. G. Leon-Saval, *et al.*, "Scaling photonic lanterns for space-division multiplexing," *Scientific Reports*, vol. 8, no. 1, p. 8897, 2018.

[26] R. Ryf, N. K. Fontaine, J. Dunayevsky, D. Sinefeld, M. Blau, M. Montoliu, S. Randel, C. Liu, B. Ercan, M. Esmaeelpour, *et al.*, "Wavelength-selective switch for few-mode fiber transmission," in *39th European Conference and Exhibition on Optical Communication (ECOC 2013)*, pp. 1–3, IET, 2013.

[27] L. Nelson, M. Feuer, K. Abedin, X. Zhou, T. Taunay, J. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, *et al.*, "Spatial superchannel routing in a two-span ROADM system for space division multiplexing," *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 783–789, 2014.

[28] M. D. Feuer, L. E. Nelson, K. Abedin, X. Zhou, T. F. Taunay, J. F. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, *et al.*, "ROADM system for space division multiplexing with spatial superchannels," in *Optical Fiber Communication Conference*, pp. PDP5B–8, Optica Publishing Group, 2013.

[29] J. Carpenter, S. G. Leon-Saval, J. R. Salazar-Gil, J. Bland-Hawthorn, G. Baxter, L. Stewart, S. Frisken, M. A. Roelens, B. J. Eggleton, and J. Schröder, "1x11 few-mode fiber wavelength selective switch using photonic lanterns," *Optics Express*, vol. 22, no. 3, pp. 2216–2221, 2014.

[30] X. Chen, A. Li, J. Ye, A. Al Amin, and W. Shieh, "Reception of mode-division multiplexed superchannel via few-mode compatible optical add/drop multiplexer," *Optics Express*, vol. 20, no. 13, pp. 14302–14307, 2012.

[31] N. K. Fontaine, R. Ryf, and D. T. Neilson, "Fiber-port-count in wavelength selective switches for space-division multiplexing," in *39th European Conference and Exhibition on Optical Communication (ECOC 2013)*, pp. 1–3, IET, 2013.

[32] M. Sharif, J. K. Perin, and J. M. Kahn, "Modulation schemes for single-laser 100 Gb/s links: Single-carrier," *Journal of Lightwave Technology*, vol. 33, no. 20, pp. 4268–4277, 2015.

[33] J. K. Perin, M. Sharif, and J. M. Kahn, "Modulation schemes for single-laser 100 Gb/s links: Multicarrier," *Journal of Lightwave Technology*, vol. 33, no. 24, pp. 5122–5132, 2015.

[34] J. Wei, J. Ingham, D. Cunningham, R. Penty, and I. White, "Comparisons between 28 Gb/s NRZ, PAM, CAP and optical OFDM systems for datacommunication applications," in *2012 Optical Interconnects Conference*, pp. 3–4, IEEE, 2012.

[35] K. Zhong, X. Zhou, T. Gui, L. Tao, Y. Gao, W. Chen, J. Man, L. Zeng, A. P. T. Lau, and C. Lu, "Experimental study of PAM-4, CAP-16, and DMT for 100 Gb/s short reach optical transmission systems," *Optics Express*, vol. 23, no. 2, pp. 1176–1189, 2015.

[36] H. Zhang, M. Li, Y. Zhang, D. Zhang, Q. Liao, J. He, S. Hu, B. Zhang, L. Wang, X. Xiao, *et al.*, "800 Gbit/s transmission over 1 km single-mode fiber using a four-channel silicon photonic transmitter," *Photonics Research*, vol. 8, no. 11, pp. 1776–1782, 2020.

[37] S. Okuda, T. Yamatoya, T. Yamaguchi, Y. Azuma, and Y. Tanaka, "High-power low-modulating-voltage 1.5 μm-band CWDM Uncooled EMLs for 800 Gb/s (53.125 Gbaud-PAM4) Transceivers," in *Optical Fiber Communication Conference*, pp. Tu1D–2, Optica Publishing Group, 2021.

[38] X. Pang, O. Ozolins, R. Lin, L. Zhang, A. Udalcovs, L. Xue, R. Schatz, U. Westergren, S. Xiao, W. Hu, *et al.*, "200 Gbps/lane IM/DD technologies for short reach optical interconnects," *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 492–503, 2020.

[39] J. Zhang, Q. Liu, M. Zhu, T. Jin, J. Zhou, H. Lin, S. Hu, X. Yi, and K. Qiu, "Beyond 200-Gb/s/λ DMT signal transmission with NGMI optimization and volterra equalization," *Journal of Lightwave Technology*, vol. 39, no. 18, pp. 5837–5844, 2021.

[40] J. K. Perin, A. Shastri, and J. M. Kahn, "Coherent data center links," *Journal of Lightwave Technology*, vol. 39, no. 3, pp. 730–741, 2020.

[41] F. Buchali, V. Aref, M. Chagnon, K. Schuh, H. Hettrich, A. Bielik, L. Altenhain, M. Guntermann, R. Schmid, and M. Moller, "1.52 Tb/s single carrier transmission supported by a 128 GSa/s SiGe DAC," in *Optical Fiber Communication Conference*, pp. Th4C–2, Optica Publishing Group, 2020.

[42] M. Mazur, J. Schröder, A. Lorences-Riesgo, T. Yoshida, M. Karlsson, and P. A. Andrekson, "11.5 bits/s/Hz PM-256QAM comb-based superchannel transmission by combining optical and digital pilots," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pp. 1–3, IEEE, 2018.

[43] S. J. Savory, "Digital signal processing options in long haul transmission," in *Optical Fiber Communication Conference*, p. OTuO3, Optica Publishing Group, 2008.

[44] P. J. Winzer, "High-spectral-efficiency optical modulation formats," *Journal of Lightwave Technology*, vol. 30, no. 24, pp. 3824–3835, 2012.

[45] M. Karlsson and E. Agrell, "Multidimensional modulation and coding in optical transport," *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 876–884, 2016.

[46] M. Eiselt, A. Dochhan, and J.-P. Elbers, “Data center interconnects at 400G and beyond,” *arXiv preprint arXiv:1807.11861*, 2018.

[47] D. S. Millar, T. Koike-Akino, S. Ö. Arık, K. Kojima, K. Parsons, T. Yoshida, and T. Sugihara, “High-dimensional modulation for coherent optical communications systems,” *Optics Express*, vol. 22, no. 7, pp. 8798–8812, 2014.

[48] T. A. Eriksson, P. Johannisson, E. Agrell, P. A. Andrekson, and M. Karlsson, “Biorthogonal modulation in 8 dimensions experimentally implemented as 2PPM-PS-QPSK,” in *Optical Fiber Communication Conference*, pp. W1A–5, Optica Publishing Group, 2014.

[49] S. J. Savory and D. S. Millar, “DSP for optical transponders,” *Springer Handbook of Optical Networks*, pp. 155–176, 2020.

[50] Z. Qu, I. B. Djordjevic, and J. Anderson, “Two-dimensional constellation shaping in fiber-optic communications,” *Applied Sciences*, vol. 9, no. 9, p. 1889, 2019.

[51] D. Pilori, A. Nespola, F. Forghieri, and G. Bosco, “Non-linear phase noise mitigation over systems using constellation shaping,” *Journal of Lightwave Technology*, vol. 37, no. 14, pp. 3475–3482, 2019.

[52] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, “Fiber-optic transmission and networking: the previous 20 and the next 20 years,” *Optics Express*, vol. 26, no. 18, pp. 24190–24239, 2018.

[53] M. P. Yankov, D. Zibar, K. J. Larsen, L. P. Christensen, and S. Forchhammer, “Constellation shaping for fiber-optic channels with QAM and high spectral efficiency,” *IEEE Photonics Technology Letters*, vol. 26, no. 23, pp. 2407–2410, 2014.

[54] R. F. Fischer, *Precoding and signal shaping for digital transmission*. John Wiley & Sons, 2005.

[55] K. Kikuchi and S. Kawakami, “Multi-level signaling in the Stokes space and its application to large-capacity optical communications,” *Optics Express*, vol. 22, no. 7, pp. 7374–7387, 2014.

[56] M. Morsy-Osman, M. Chagnon, M. Poulin, S. Lessard, and D. V. Plant, “224-Gb/s 10-km transmission of PDM PAM-4 at 1.3 μm using a single intensity-modulated laser and a direct-detection MIMO DSP-based receiver,” *Journal of Lightwave Technology*, vol. 33, no. 7, pp. 1417–1424, 2015.

[57] M. Chagnon, M. Morsy-Osman, D. Patel, V. Veerasubramanian, A. Samani, and D. Plant, “Digital signal processing for dual-polarization intensity and interpolarization phase modulation formats using stokes detection,” *Journal of Lightwave Technology*, vol. 34, no. 1, pp. 188–195, 2015.

[58] M. Chagnon and D. V. Plant, “504 and 462 Gb/s direct detect transceiver for single carrier short-reach data center applications,” in *Optical Fiber Communication Conference*, pp. W3B–2, Optica Publishing Group, 2017.

[59] M. Morsy-Osman, M. Chagnon, M. Poulin, S. Lessard, and D. V. Plant, “1$\lambda\times$ 224 Gb/s 10 km transmission of polarization division multiplexed PAM-4 signals using 1.3 μm SiP intensity modulator and a direct-detection MIMO-based receiver,” in *2014 The European Conference on Optical Communication (ECOC)*, pp. 1–3, IEEE, 2014.

[60] M. Morsy-Osman, M. Chagnon, and D. V. Plant, “Four-dimensional modulation and Stokes direct detection of polarization division multiplexed intensities, inter polarization phase and inter polarization differential phase,” *Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1585–1592, 2016.

[61] M. Y. Sowailem, T. M. Hoang, M. Chagnon, M. Morsy-Osman, M. Qiu, S. Paquet, C. Paquet, I. Woods, O. Liboiron-Ladouceur, and D. V. Plant, “100G and 200G single carrier transmission over 2880 and 320 km using an InP IQ modulator and Stokes vector receiver,” *Optics Express*, vol. 24, no. 26, pp. 30485–30493, 2016.

[62] A. Li, Z. Li, Y. Wen, W.-R. Peng, Y. Cui, and Y. Bai, “192-Gb/s 160-km transmission of carrier-assisted dual-polarization signal with Stokes vector direct detection,” in *Optical Fiber Communication Conference*, pp. W1A–2, Optica Publishing Group, 2016.

[63] W. O. Popoola, E. Poves, and H. Haas, “Spatial pulse position modulation for optical communications,” *Journal of Lightwave Technology*, vol. 30, no. 18, pp. 2948–2954, 2012.

[64] J. Van Weerdenburg, A. Alvarado, J. A. Zacarias, J. A. Lopez, J. Bonarius, D. Molin, M. Bigot-Astruc, A. M. Koonen, A. A. Correa, P. Sillard, *et al.*, “Spatial pulse position modulation for multi-mode transmission systems,” in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, IEEE, 2017.

[65] T.-C. Bui, M. Biagi, and S. Kiravittaya, “Theoretical analysis of optical spatial multiple pulse position modulation,” in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, IEEE, 2018.

[66] D. V. Plant, M. H. Morsy-Osman, M. Chagnon, and S. Lessard, “Trends in high-speed interconnects for datacenter networking: Multidimensional formats and their enabling DSP,” *Optical Switching in Next Generation Data Centers*, pp. 261–277, 2018.

[67] D. Che, A. Li, X. Chen, Q. Hu, Y. Wang, and W. Shieh, “Stokes vector direct detection for linear complex optical channels,” *Journal of Lightwave Technology*, vol. 33, no. 3, pp. 678–684, 2014.

[68] H. Ji, D. Che, C. Sun, J. Fang, G. Milione, R. R. Unnithan, and W. Shieh, “High-dimensional Stokes vector direct detection over few-mode fibers,” *Optics Letters*, vol. 44, no. 8, pp. 2065–2068, 2019.

[69] D. Che and W. Shieh, "Polarization demultiplexing for Stokes vector direct detection," *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 754–760, 2016.

[70] I. Roudas, J. Kwapisz, and E. Fink, "Mode vector modulation," in *2021 European Conference on Optical Communication (ECOC)*, pp. 1–4, IEEE, 2021.

[71] J. Kwapisz, I. Roudas, and E. Fink, "Mode Vector Modulation: Extending Stokes Vector Modulation to Higher Dimensions," *Journal of Lightwave Technology*, 2023.

[72] T. Hoang, M. Sowailem, Q. Zhuge, M. Osman, A. Samani, C. Paquet, S. Paquet, I. Woods, and D. Plant, "Enabling high-capacity long-reach direct detection transmission with QAM-PAM Stokes vector modulation," *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 460–467, 2017.

[73] M. V. Bnyamin, X. Jiang, and M. D. Feuer, "Independent component analysis for impairment mitigation in direct-detected Stokes vector modulation," *Optics Express*, vol. 30, no. 25, pp. 45350–45364, 2022.

[74] N. Cui, Z. Zheng, X. Zhang, W. Yi, R. Guo, W. Zhang, X. Tang, H. Xu, and L. Xi, "Joint blind equalization of CD and RSOP using a time-frequency domain Kalman filter structure in Stokes vector direct detection system," *Optics Express*, vol. 27, no. 8, pp. 11557–11570, 2019.

[75] L. Cao, H. Xu, Z. Sun, C. Bai, N. Cui, Y. Zhang, L. Yang, W. Sun, and Y. Bi, "Joint equalization of EEPN, RSOP, and CD with the sliding window assisted extended Kalman filter for a high baud rate Stokes vector direct detection system," *Applied Optics*, vol. 62, no. 4, pp. 1066–1075, 2023.

[76] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[77] E. Oja and A. Hyvarinen, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[78] Z. He, X. Li, M. Luo, R. Hu, C. Li, Y. Qiu, S. Fu, Q. Yang, and S. Yu, "Independent component analysis based channel equalization for 6× 6 MIMO-OFDM transmission over few-mode fiber," *Optics Express*, vol. 24, no. 9, pp. 9209–9217, 2016.

[79] A. Nafta, P. Johannisson, and M. Shtaif, "Blind equalization in optical communications using independent component analysis," *Journal of Lightwave Technology*, vol. 31, no. 12, pp. 2043–2049, 2013.

[80] Q. Xiang, Y. Yang, J. Cao, Q. Zhang, and Y. Yao, "Polarization tracking for Stokes-vector modulation formats using Kalman filter," in *2018 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, pp. 1–2, IEEE, 2018.

[81] D. Li, H. Song, W. Cheng, M. Cheng, S. Fu, M. Tang, D. Liu, and L. Deng, "Low-complexity equalization scheme for suppressing FFE-enhanced in-band noise and ISI in 100 Gbps PAM4 optical IMDD system," *Optics Letters*, vol. 45, no. 9, pp. 2555–2558, 2020.

[82] X. Li, X. Zhang, Z. Zheng, N. Cui, N. Zhang, W. Zhang, X. Tang, H. Zhang, and L. Xi, "Polarization demultiplexing scheme for probabilistic shaping Stokes vector direct detection system using extended Kalman filter," *Optics Communications*, vol. 461, p. 125192, 2020.

[83] I. Roudas, "Modal dispersion characterization of multimode fibers," in *2017 IEEE Photonics Conference (IPC)*, pp. 411–412, IEEE, 2017.

[84] I. Roudas and J. Kwapisz, "Accurate modal dispersion measurements using maximally-orthogonal Stokes vectors," in *2018 Conference on Lasers and Electro-Optics (CLEO)*, pp. 1–2, IEEE, 2018.

[85] T. Joseph and J. John, "Exciting single modes (LP01 or LP11) in MMFs for High Speed Applications," in *International Conference on Fibre Optics and Photonics*, pp. Tu3A–4, Optica Publishing Group, 2016.

[86] C. Antonelli, A. Mecozzi, and M. Shtaif, "The delay spread in fibers for SDM transmission: dependence on fiber parameters and perturbations," *Optics Express*, vol. 23, no. 3, pp. 2196–2202, 2015.

[87] I. Roudas and J. Kwapisz, "Optimization of the Mode-Dependent Signal Delay Method for the Measurement of Modal Dispersion," in *2018 IEEE Photonics Society Summer Topical Meeting Series (SUM)*, pp. 225–226, IEEE, 2018.

[88] J. M. Gené and P. J. Winzer, "A universal specification for multicore fiber crosstalk," *IEEE Photonics Technology Letters*, vol. 31, no. 9, pp. 673–676, 2019.

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

[90] M. Hayes, *Statistical Digital Signal Processing and Modeling*. Wiley India Pvt. Limited, 2009.

[91] G. Bishop, G. Welch, *et al.*, "An introduction to the Kalman filter," *Proc of SIGGRAPH, Course*, vol. 8, no. 27599-23175, p. 41, 2001.

[92] T. Marshall, B. Szafraniec, and B. Nebendahl, "Kalman filter carrier and polarization-state tracking," *Optics Letters*, vol. 35, no. 13, pp. 2203–2205, 2010.

[93] N. J. Muga and A. N. Pinto, "Extended Kalman filter vs. geometrical approach for Stokes space-based polarization demultiplexing," *Journal of Lightwave Technology*, vol. 33, no. 23, pp. 4826–4833, 2015.

[94] J. Yang and D. A. Nolan, "Using state tomography for characterizing input principal modes in optically scattering medium," *Optics Express*, vol. 24, no. 24, pp. 27691–27701, 2016.

[95] Bandyopadhyay, Boykin, Roychowdhury, and Vatan, "A new proof for the existence of mutually unbiased bases," *Algorithmica*, vol. 34, pp. 512–528, 2002.

[96] C. A. Fuchs, M. C. Hoang, and B. C. Stacey, "The SIC question: History and state of play," *Axioms*, vol. 6, no. 3, p. 21, 2017.

[97] J. M. Kahn, K.-P. Ho, and M. B. Shemirani, "Mode coupling effects in multi-mode fibers," in *Optical Fiber Communication Conference*, pp. OW3D–3, Optica Publishing Group, 2012.