Demodulation and Detection Schemes for a Memoryless Optical WDM Channel

(article starts on next page)

Demodulation and Detection Schemes for a Memoryless Optical WDM Channel

Kamran Keykhosravi, *Student Member, IEEE,* Morteza Tavana, *Student Member, IEEE,* Erik Agrell, *Fellow, IEEE,* and Giuseppe Durisi, *Senior Member, IEEE.*

Abstract—It is well known that matched filtering and sampling (MFS) demodulation together with minimum Euclidean distance (MD) detection constitute the optimal receiver for the additive white Gaussian noise channel. However, for a general nonlinear transmission medium, MFS does not provide sufficient statistics, and therefore is suboptimal. Nonetheless, this receiver is widely used in optical systems, where the Kerr nonlinearity is the dominant impairment at high powers. In this paper, we consider a suite of receivers for a two-user channel subject to a type of nonlinear interference that occurs in wavelength-division-multiplexed channels. The asymptotes of the symbol error rate (SER) of the considered receivers at high powers are derived or bounded analytically. Moreover, Monte-Carlo simulations are conducted to evaluate the SER for all the receivers. Our results show that receivers that are based on MFS cannot achieve arbitrary low SERs, whereas the SER goes to zero as the power grows for the optimal receiver. Furthermore, we devise a heuristic demodulator, which together with the MD detector yields a receiver that is simpler than the optimal one and can achieve arbitrary low SERs. The SER performance of the proposed receivers is also evaluated for some single-span fiber-optical channels via split-step Fourier simulations.

Index Terms—Optical fiber, nonlinearity compensation, nonlinear channel, demodulation, MAP detector.

I. Introduction

The development of the standard single-mode fiber (SMF) in the 1970s [2] and of the erbium-doped fiber amplifiers [3] in the late 80s increased the capacity of the fiber-optical channel far beyond the required data rate in those days. This abundance of resources made it inessential to exploit the bandwidth optimally in the design of optical communication networks. Nowadays, however, with the exponential growth of the global Internet, the data demand has started meeting the limits of traditional optical systems. This ever-increasing data demand has motivated many recent efforts, including the one in the current paper, to increase the efficiency of optical transmitters and receivers.

For the additive white Gaussian noise (AWGN) channel, it is well known that the matched filtering and sampling (MFS) demodulator provides sufficient statistics for detecting the transmitted symbol from the received continuous-time signal. Although, in general, this method is suboptimal for nonlinear channels, it has been deployed broadly in optical fiber transmission systems, where the Kerr nonlinearity critically limits the achievable information rate at moderate and high powers [4].

In advanced optical communication systems, a single-mode fiber hosts approximately one hundred wavelength-division-multiplexed (WDM) channels. In such systems, the Kerr nonlinearity gives rise to: *i)* self-phase modulation (SPM), where the signal phase is distorted depending on its own magnitude; *ii)* cross-phase modulation (XPM), where the magnitude of the signal transmitted over neighboring channels modulates the phase of the signal of interest; and *iii)* four-wave mixing (FWM), where three signals at different frequencies create a distortion at a new frequency. In this paper, we shall focus on the first two effects and assume that the impact of FWM (the third effect) is mitigated by appropriate channel spacing (see, for example, [5]).

Many methods have been proposed, both in the optical and the electrical domains, to compensate for the fiber nonlinear distortion [6, Ch. 2]. Soliton-based communication [7] is among the primary solutions to mitigate the channel impairments including the nonlinearity. It is based on soliton pulses, which can propagate through the fiber undisturbed. In recent years, this method has received attention in the context of the nonlinear Fourier transform [8]. Inverting the signal's phase at the middle of the transmission line is another effective approach to reduce the nonlinear distortion [9].

In the last decade, the advancement of digital signal processors (DSP) made them a key enabling technology for data transmission over the fiber-optical channel. A number of known nonlinearity mitigation techniques are based on DSPs, three of which are reviewed next. *i) Digital back propagation* [10] is a well-known method to compensate for the fiber impairments. Using this technique, all the signal–signal distortions can be compensated for by processing the signal at the transmitter, at the receiver, or at both

ends. However, digital back propagation suffers from high computational complexity, and it requires knowledge of all copropagating channels. *ii)* The effects of XPM can be partially mitigated via *adaptive equalization* that utilizes the time coherency of the XPM distortions (see for example [11], [12]). *iii)* Using an approximate probability distribution for the channel law, one can devise *nonlinearity-tailored detection* techniques to improve the symbol error rate (SER) [13], [14].

The optical channel model can be described by the nonlinear Schrödinger (NLS) equation [15, Eq. 2.6.18]. Since the input–output relation is given implicitly through a differential equation, developing the optimal transmitter and receiver for the NLS channel seems a formidable task. By neglecting the channel memory, closed-form input–output relations can be obtained. The analyses based on these models are applicable to optical systems with short-haul zero-dispersion fibers (see, for example [16]–[20]). Furthermore, many simplified models have been developed in the literature to approximate the NLS channel (see [21] and the references therein). Applying perturbation theory, or equivalently Volterra series, and ignoring signal–noise interaction are among the most common simplifications. The channel models derived based on these assumptions lose accuracy at high powers [22]. Nevertheless, since the physical channel is intractable, these models can be studied to develop transceivers that are more matched to the nonlinear nature of the optical channel than the MFS. The corresponding results can serve as a first step towards optimizing optical receivers for the actual physical channel.

In [23], [24], the capacity of a memoryless discrete-time two-user WDM channel, where both SPM and XPM are present, has been studied at high powers. It has been proved for this channel that the capacity pre-log[1] pair $(1,1)$ is achievable. The discrete-time channel model used in [23], [24] relies on the sampling receiver, whose bandwidth is infinite. This receiver has been used in many publications to obtain a tractable discrete-time model for the single-user NLS channel (see, for example [19], [20], [22], [25]). However, the sampling receiver is suboptimal and impractical, particularly for WDM systems [26, Sec. I]. The discrete-time channel in [23], [24] can also be obtained from the underlying continuous-time channel by using rectangular pulse shaping at the modulator, which, however, cannot be implemented in practice.

This paper studies the same continuous-time two-user WDM channel as in [23], [24]. Although our focus in this paper is on a two-user channel, our framework can be used to analyze a channel of interest in a WDM system with an arbitrary number of users by considering all of the interfering signals as a single channel [27], [28]. We consider three demodulation schemes for the aforementioned continuous-time channel under the assumption that joint processing is not possible at the transmitters or at the receivers. First, the MFS demodulator is studied, which is conventionally used in optical systems. Second, a demodulator that provides sufficient statistics (SS) is developed. Third, a novel heuristic demodulation method, referred to as maximum matching (MxM), is presented. Furthermore, three different detection schemes, used at the receivers to estimate the transmitted signal based on the demodulator output, are considered: the conventional minimum Euclidean distance (MD) detector, the optimal detector based on maximum a posteriori (MAP) probability, and a two-stage (TS) detection method, which first estimates the amplitude and then the phase of each symbol. Different versions of TS detectors have been considered previously to mitigate the nonlinear phase noise in optical systems [17], [25], [29]. As we shall see, our TS detector is superior to the MD detector at moderate powers.

By coupling different modulators and detectors, we investigate the performance (in terms of SER) and the complexity of six different receivers. First, we study the conventional MFS-MD receiver, which is optimal for the linear AWGN channel. Second, we study a receiver that performs MFS demodulation, phase recovery (using the method in [30]), and MD detection. This receiver, which relies on processing techniques used in today's optical systems, is referred to as MFS-PR. Third, to find the performance limits of the MFS demodulator, we couple it with the optimal (MAP) detector. Fourth, we consider the SS-MAP receiver, which is the optimal receiver for the channel under study. Fifth, we couple MxM with MD to obtain a receiver that has a lower complexity than SS-MAP and can achieve arbitrary low SERs. Sixth, we study the MxM-TS receiver, which turns out to yield a slight performance improvement over MxM-MD at moderate powers. A summary of the considered receivers and a qualitative evaluation of their complexity and performance is provided in Table I. At low powers, where nonlinearity is weak, all the receivers except the MxM-TS and MFS-PR have approximately the same SER as the optimal receiver, whereas in the moderate-power regime only MFS-MAP and MFS-PR perform close to optimal. It can be seen that unlike receivers based on MFS, the SER for the optimal receiver (SS-MAP) goes to zero as the power grows large. Also, arbitrarily low SERs can be achieved via simple detectors (MD and TS) coupled with the MxM demodulator. The results presented in Table I are obtained for truncated Gaussian pulse shaping and 16-QAM modulation. We expect similar results to hold for practically relevant pulse shapes whose spectrum broadens with increasing power (see [19, Sec. VIII]). For rectangular pulse shaping, for which the signal spectrum does not broaden, MFS provides sufficient statistics and the SER of MFS-MAP goes to zero as power grows large. Modulation formats that are resilient to phase noise, such as pulse-amplitude modulation, may also result in a different SNR behavior compared to Table I.

We also evaluate the SER performance of the proposed receivers (by means of split-step Fourier simulations) for two single-span fiber-optical systems with different disper-

---

[1]The capacity pre-log is defined by $\lim_{P \to \infty} \mathcal{C}(P)/\log P$, where $\mathcal{C}(P)$ is the channel capacity under the input power constraint P.

TABLE I
A QUALITATIVE COMPARISON BETWEEN THE COMPLEXITY AND PERFORMANCE OF THE RECEIVERS UNDER STUDY.

Receiver	Complexity		Symbol error rate compared to the optimal receiver		
	Demodulation	Detection	Low powers	Moderate powers	High powers
MFS-MD	Low	Low	Close to optimal	Far from optimal	Far from optimal
MFS-PR	Low	Low	Far from optimal	Close to optimal	Far from optimal
MFS-MAP	Low	High	Close to optimal	Close to optimal	Far from optimal
SS-MAP	High	High	Optimal	Optimal	Optimal $(\rightarrow 0)^*$
MxM-MD	High	Low	Close to optimal	Far from optimal	Close to optimal $(\rightarrow 0)$
MxM-TS	High	Low	Far from optimal	Far from optimal†	Close to optimal $(\rightarrow 0)$

* SER$\rightarrow 0$ as power grows large. † The SER with MxM-TS, is lower than with MxM-MD at moderate powers.

sion parameters. Our results show that, for all receivers, the SER increases with power after a certain optimal power. When the dispersion is small, the performance of SS-MAP and MFS-MAP turns out to be superior to that of MFS-PR. When dispersion is high, all receivers except for SS-MAP are inferior to MFS-PR. This paper completes the analysis initiated with the conference paper [1], where the MxM-MD and the MFS-MD receivers were investigated for the channel under study.

Organization: The rest of this paper is organized as follows. In Section II, a model for a continuous-time two-user WDM channel is obtained from a pair of coupled NLS equations under some simplifications. In Section III, we present the demodulation and detection methods. Section IV presents some analytical asymptotic bounds on the SER. Numerical results are provided in Section V. Specifically, in Section V-A, we study the simplified channel model and in Section V-B, the performance under more realistic dispersive conditions is evaluated by simulation. Finally, Section VI concludes the paper.

Notation: Bold-face letters are used to denote random quantities. Sets are indicated by upper-case script letters, e.g., $\mathcal{X}$. The cardinality of a set $\mathcal{X}$ is indicated by $|\mathcal{X}|$. Vectors are denoted by lower-case underlined letters. $\mathcal{CN}(\mu, \sigma^2)$ denotes the proper complex Gaussian distribution with mean μ and variance σ^2. The inner product between two complex functions $f(t)$ and $g(t)$ is defined as $\langle f, g \rangle = \int_{-\infty}^{\infty} f(t)g^*(t)\, \mathrm{d}t$, where $(\cdot)^*$ denotes complex conjugation. $\Re(x)$ and $\Im(x)$ denote the real and the imaginary part of a complex number x, respectively. With $|\cdot|$ and $(\cdot)^T$ we denote the determinant and the transpose operators, respectively. We use $\Pr(\mathbf{x} = x)$ to denote the probability mass function of a discrete random variable $\mathbf{x}$ at x. Also, the probability density function of a continuous random variable $\mathbf{x}$ at x is denoted by $f_{\mathbf{x}}(x)$. The real line and the complex plane are represented by $\mathbb{R}$ and $\mathbb{C}$, respectively. Finally, for two functions $q(x)$ and $r(x)$, we write $q(x) = \mathcal{O}(r(x))$ if $\limsup_{x \to 0} |q(x)/r(x)| < \infty$.

II. CHANNEL MODEL

The signal propagation through the fiber-optical channel suffers from several impairments such as chromatic dispersion, fiber loss, and Kerr nonlinearity. The chromatic dispersion is mainly caused by the dependency of the refractive index on the frequency. Therefore, in the presence of chromatic dispersion, the different frequency components of a transmitted pulse propagate with different speeds, causing the pulse to broaden in time. This impairment can be compensated for by using dispersion-compensating fibers or through DSPs.

To compensate for the fiber loss, two types of optical amplification are typically deployed, namely, distributed or lumped amplification. While the former amplifies the signal continuously during propagation, the latter does so only at the end of each amplification span. Optical amplification is always accompanied by additive noise caused by spontaneously emitted light photons. In this paper, we shall focus on lumped-amplified systems.

The main impairment that limits the achievable data rates in fiber communications is the Kerr nonlinearity. It arises because the glass refractive index depends on the propagating optical power. It can be described by a phase shift proportional to the optical power applied to the complex baseband signal. This phase shift is caused by the signal itself (SPM) or by other copropagating signals at different wavelengths (XPM).

In this paper, we consider two channel models: a simple memoryless model for algorithm design and analysis, and a more realistic split-step Fourier model for performance evaluation. For the first purpose, we consider the propagation of two optical signals with different carrier wavelengths through a point-to-point single-mode fiber, focusing on the effects of SPM and XPM. We assume that the two signals have nonoverlapping spectra. The signal propagation can then be described by the pair of coupled NLS equations [31, Eqs. (7.4.1)–(7.4.2)]

$$\frac{\partial \mathbf{a}_1}{\partial z} + \frac{j\beta_{21}}{2}\frac{\partial \mathbf{a}_1}{\partial t^2} + \frac{\alpha}{2}\mathbf{a}_1 = j\gamma_1 \left(|\mathbf{a}_1|^2 + 2|\mathbf{a}_2|^2 \right) \mathbf{a}_1 \tag{1}$$

$$\frac{\partial \mathbf{a}_2}{\partial z} + d\frac{\partial \mathbf{a}_2}{\partial t} + \frac{j\beta_{22}}{2}\frac{\partial \mathbf{a}_2}{\partial t^2} + \frac{\alpha}{2}\mathbf{a}_2 = j\gamma_2 \left(|\mathbf{a}_2|^2 + 2|\mathbf{a}_1|^2 \right) \mathbf{a}_2 \tag{2}$$

where $\mathbf{a}_k = \mathbf{a}_k(z,t)$, $k \in \{1, 2\}$ is the complex envelope of the optical signal k at position z and time t. Time is measured according to a reference frame moving with

$\mathbf{a}_1(z,t)$. The group-velocity mismatch between the two channels is given by d. The constants β_{2k} and γ_k are the dispersion and the nonlinearity coefficients, respectively. The fiber loss, which is assumed to be the same in both channels, is quantified by the parameter α. Although our focus in this paper is on single-polarization transmission, our analytical framework can be adapted to suit an extension of the channel model (1)–(2) to two polarizations (see [32, Eqs. (7.1.19)–(7.1.20)]).

We assume that the fiber loss is completely compensated for using lumped amplification and that each amplifier generates Gaussian noise. Moreover, we assume that the effects of dispersion, group velocity mismatch, and signal–noise interaction are negligible. This assumption is valid for single-span short-haul communication systems with (optical or digital) dispersion compensation. Under this assumption, the coupled NLS equations (1)–(2) yield the continuous-time channel [31, Eq. (7.4.5)]

$$\mathbf{a}_1(L,t) = \mathbf{a}_1(0,t)e^{j\eta_1\left(|\mathbf{a}_1(0,t)|^2 + 2|\mathbf{a}_2(0,t)|^2\right)} + \mathbf{n}_1(t) \quad (3)$$

$$\mathbf{a}_2(L,t) = \mathbf{a}_2(0,t)e^{j\eta_2\left(|\mathbf{a}_2(0,t)|^2 + 2|\mathbf{a}_1(0,t)|^2\right)} + \mathbf{n}_2(t). \quad (4)$$

Here, L is the length of the fiber. The parameters η_k quantify the nonlinearity and can be calculated as

$$\eta_k = n_{\text{span}}\gamma_k L_{\text{eff}} \quad (5)$$

where n_{span} is the number of amplification spans and

$$L_{\text{eff}} = \frac{1 - e^{-\alpha L_{\text{span}}}}{\alpha} \quad (6)$$

is the effective length of the fiber in a single span with length $L_{\text{span}} = L/n_{\text{span}}$. Because of fiber loss, the signal power and, consequently, the nonlinear distortion, diminishes along the fiber. Therefore, the effective length is less than the actual span length L_{span}. Finally, the amplification noise is captured by $\mathbf{n}_1(t)$ and $\mathbf{n}_2(t)$, which are two independent complex white circularly-symmetric Gaussian processes with power spectral density

$$N_0 = \frac{1}{2}n_{\text{span}}h\nu FG. \quad (7)$$

Here, $h\nu$ is the optical photon energy, F is the noise figure, and G is the amplifier gain, which we assume equal to the signal attenuation in one span $\exp(\alpha L_{\text{span}})$.

In this paper, we shall first focus on the simplified continuous-time model (3)–(4) and study the SER performance of different demodulation and detection schemes. A more realistic channel model is studied in Section V-B. Throughout the paper, we assume that the parameters of the fiber are known at both receivers. Moreover, we assume that the messages sent over each channel are independent, and that joint processing is not allowed at the transmitters or receivers.

III. Modulation, Demodulation, and Detection

In this section, a modulation scheme together with the six receivers listed in Table I are presented for the continuous-time channel (3)–(4). The transmitters are assumed to perform linear modulation. Specifically, let the pulse shape $g(t)$ be a real function that is zero outside the interval $(0, T]$ and has unit energy, i.e., $\int_0^T g^2(t)\,dt = 1$. Furthermore, define $\mathbf{a}_k(0,t) = \sum_i \mathbf{x}_{ki}g(t - iT)$ to be the signal sent by transmitter k, where $\mathbf{x}_{ki} \in \mathbb{C}$ is the ith transmitted symbol. Since $g(t)$ is zero outside $(0, T]$, after demodulation the noise terms at different symbol times become independent. Based on this and the fact that the channel model is memoryless, the channel can be studied by only considering the input–output relation in the first symbol interval. Hence, we can drop the index i. For $0 \le t < T$, the channel (3) can be expressed as

$$\mathbf{a}_1(L,t) = \mathbf{x}_1 g(t)\exp\left(j\eta_1\left(|\mathbf{x}_1|^2 + 2|\mathbf{x}_2|^2\right)g^2(t)\right) + \mathbf{n}_1(t) \quad (8)$$

where we set $\mathbf{x}_k = \mathbf{x}_{k1}$ for $k = 1, 2$ to simplify notation. In this section, we focus only on the first WDM channel (3). Because of the symmetry, all the results hold for the second channel (4) as well.

Next, we introduce some notation that will come to use in the rest of this section. We assume that the input random variable $\mathbf{x}_1$ takes values from a finite-cardinality set $\mathcal{X} = \{x_1, x_2, ..., x_{|\mathcal{X}|}\}$ and has a probability distribution $\pi_i = \Pr(\mathbf{x}_1 = x_i)$. Furthermore, we assume that $\mathbf{x}_2$ belongs to a finite-cardinality set, which may be different from $\mathcal{X}$. Also, we let $\mathbf{s} = |\mathbf{x}_1|^2 + 2|\mathbf{x}_2|^2$, which belongs to a finite set $\mathcal{S} = \{s_1, s_2, ..., s_{|\mathcal{S}|}\}$. Finally, we denote the conditional probability distribution of $\mathbf{s}$ given $\mathbf{x}_1$ by $\tilde{\pi}_{ji} = \Pr(\mathbf{s} = s_j \mid |\mathbf{x}_1| = |x_i|)$.

Next, we study the receivers listed in Table I. We begin by introducing the conventional MFS-MD and MFS-PR receivers. Then, we study MFS-MAP, which is used to determine the performance limits of MFS demodulation. Next, we devise the optimal receiver, SS-MAP, which serves as a benchmark to assess the performance of the other receivers. Finally, two heuristic receivers, MxM-MD and MxM-TS are studied. These receivers have lower complexity than SS-MAP and can obtain arbitrary low SERs for sufficiently high powers.

A. MFS demodulation with MD detection (MFS-MD)

The MFS demodulator maps the received signal $\mathbf{a}_1(L,t)$ to the complex number

$$\mathbf{v} = \int_0^T \mathbf{a}_1(L,t) \cdot g(t)\,dt \quad (9)$$

$$= \langle \mathbf{a}_1(L,t), g(t)\rangle. \quad (10)$$

After observing the demodulation outcome $\mathbf{v} = v$, the MD detector selects $x_m \in \mathcal{X}$ such that

$$m = \arg\min_i |v - x_i|^2. \quad (11)$$

B. MFS demodulation with phase recovery (MFS-PR)

In the MFS-PR receiver, the output of the MFS demodulator passes through a phase-recovery block and is then

fed to the MD detector. Throughout, we shall focus on the phase-recovery technique proposed in [30].[2]

C. MFS demodulation with MAP detection (MFS-MAP)

Given the MFS output $\mathbf{v} = v$ in (9), the optimal MAP detector determines the input symbol $x_m \in \mathcal{X}$, such that

$$m = \arg\max_i \Pr(\mathbf{x}_1 = x_i \mid \mathbf{v} = v) \tag{12}$$

$$= \arg\max_i \pi_i f_{\mathbf{v}|\mathbf{x}_1}(v \mid x_i) \tag{13}$$

$$= \arg\max_i \pi_i \sum_j \tilde{\pi}_{ji} f_{\mathbf{v}|\mathbf{s},\mathbf{x}_1}(v \mid s_j, x_i) \tag{14}$$

where in (14) we used that $\Pr(\mathbf{s} = s_j \mid \mathbf{x}_1 = x_i) = \Pr(\mathbf{s} = s_j \mid |\mathbf{x}_1| = |x_i|) = \tilde{\pi}_{ji}$. The conditional probability $f_{\mathbf{v}|\mathbf{s},\mathbf{x}_1}(v \mid s_j, x_i)$ can be calculated by noting that, given $\mathbf{s} = s_j$ and $\mathbf{x}_1 = x_i$, we have that $\mathbf{v} \sim \mathcal{CN}(\mu_{ji}, N_0)$, where

$$\mu_{ji} = x_i \left\langle g(t) \exp\left(j\eta_1 s_j g^2(t)\right), g(t) \right\rangle. \tag{15}$$

Therefore,

$$f_{\mathbf{v}|\mathbf{s},\mathbf{x}_1}(v \mid s_j, x_i) = \frac{1}{\pi N_0} \exp\left(-\frac{|v - \mu_{ji}|^2}{N_0}\right). \tag{16}$$

D. Sufficient statistics with MAP detection (SS-MAP)

Let $\phi(\mathbf{s}, t) = \eta_1 \, \mathbf{s} \, g^2(t)$. The real and the imaginary part of $\mathbf{a}_1(L, t)$ are

$$\Re(\mathbf{a}_1(L, t)) = \Re(\mathbf{x}_1) \, g(t) \cos(\phi(\mathbf{s}, t))$$
$$- \Im(\mathbf{x}_1) \, g(t) \sin(\phi(\mathbf{s}, t)) + \Re(\mathbf{n}_1(t)) \tag{17}$$

$$\Im(\mathbf{a}_1(L, t)) = \Re(\mathbf{x}_1) \, g(t) \sin(\phi(\mathbf{s}, t))$$
$$+ \Im(\mathbf{x}_1) \, g(t) \cos(\phi(\mathbf{s}, t)) + \Im(\mathbf{n}_1(t)). \tag{18}$$

Note that, if additive noise is neglected, the signals $\Re(\mathbf{a}_1(L, t))$ and $\Im(\mathbf{a}_1(L, t))$ can be written as linear combinations of the signals $h_\ell(t) = g(t) \sin(\phi(s_\ell, t))$ and $\tilde{h}_\ell(t) = g(t) \cos(\phi(s_\ell, t))$, $\ell = 1, \ldots, |\mathcal{S}|$. Therefore, by [33, Corollary 26.4.2],

$$\mathbf{u}_\ell^{\mathrm{R}} = \langle \Re(\mathbf{a}_1(L, t)), h_\ell(t) \rangle \tag{19}$$

$$\tilde{\mathbf{u}}_\ell^{\mathrm{R}} = \langle \Re(\mathbf{a}_1(L, t)), \tilde{h}_\ell(t) \rangle \tag{20}$$

$$\mathbf{u}_\ell^{\mathrm{I}} = \langle \Im(\mathbf{a}_1(L, t)), h_\ell(t) \rangle \tag{21}$$

$$\tilde{\mathbf{u}}_\ell^{\mathrm{I}} = \langle \Im(\mathbf{a}_1(L, t)), \tilde{h}_\ell(t) \rangle \tag{22}$$

are sufficient statistics for determining $\mathbf{x}_1$ based on $\mathbf{a}_1(L, t)$. Let $\underline{\mathbf{u}}^{\mathrm{R}} = [\mathbf{u}_1^{\mathrm{R}}, \ldots, \mathbf{u}_{|\mathcal{S}|}^{\mathrm{R}}]$, and similarly define the vectors $\tilde{\underline{\mathbf{u}}}^{\mathrm{R}}$, $\underline{\mathbf{u}}^{\mathrm{I}}$, and $\tilde{\underline{\mathbf{u}}}^{\mathrm{I}}$. Moreover, let the vector $\underline{\mathbf{u}}$ with length $4|\mathcal{S}|$ be the concatenation of the aforementioned vectors, i.e.,

$$\underline{\mathbf{u}} = [\underline{\mathbf{u}}^{\mathrm{R}}, \tilde{\underline{\mathbf{u}}}^{\mathrm{R}}, \underline{\mathbf{u}}^{\mathrm{I}}, \tilde{\underline{\mathbf{u}}}^{\mathrm{I}}]. \tag{23}$$

It follows from [33, Prop. 25.15.2] that the vector $\underline{\mathbf{u}}$ is conditionally jointly Gaussian given $\mathbf{s} = s_j$ and $\mathbf{x}_1 = x_i$. Let the conditional mean vector of $\underline{\mathbf{u}}$ given $\mathbf{s} = s_j$ and $\mathbf{x}_1 = x_i$ be $\underline{\mu}_{ji}$ and the conditional covariance matrix be Σ (as we shall see later, Σ does not depend on j or i).

---

[2]The test carrier phases considered in the simulation results are $\pi b/128$, $b \in \{-32, \ldots, 31\}$.

It follows from steps similar to (12)–(14) that the MAP decoder, after observing $\underline{\mathbf{u}} = u$, selects the transmitted symbol x_m such that

$$m = \arg\max_i \pi_i \sum_j \tilde{\pi}_{ji} f_{\underline{\mathbf{u}}|\mathbf{s},\mathbf{x}_1}(u \mid s_j, x_i) \tag{24}$$

where

$$f_{\underline{\mathbf{u}}|\mathbf{s},\mathbf{x}_1}(u \mid s_j, x_i) = \frac{\exp\left(-\frac{1}{2}(u - \underline{\mu}_{ji})\Sigma^{-1}(u - \underline{\mu}_{ji})^T\right)}{(2\pi)^{2|\mathcal{S}|}\sqrt{|\Sigma|}}. \tag{25}$$

Next we calculate $\underline{\mu}_{ji}$ and Σ. We write $\underline{\mu}_{ji}$ as a concatenation of four vectors: $\underline{\mu}_{ji} = [\underline{\mu}_{ji}^{\mathrm{R}}, \tilde{\underline{\mu}}_{ji}^{\mathrm{R}}, \underline{\mu}_{ji}^{\mathrm{I}}, \tilde{\underline{\mu}}_{ji}^{\mathrm{I}}]$. It follows from (23) that the ℓth element of $\underline{\mu}_{ji}^{\mathrm{R}}$ is

$$\mu_{ji\ell}^{\mathrm{R}} = \mathrm{E}\left[\mathbf{u}_\ell^{\mathrm{R}} \mid \mathbf{s} = s_j, \mathbf{x}_1 = x_i\right]. \tag{26}$$

The vectors $\tilde{\underline{\mu}}_{ji}^{\mathrm{R}}$, $\underline{\mu}_{ji}^{\mathrm{I}}$, and $\tilde{\underline{\mu}}_{ji}^{\mathrm{I}}$ can be calculated as in (26). We have from (17) and (19) that

$$\mathrm{E}\left[\mathbf{u}_\ell^{\mathrm{R}} \mid \mathbf{s} = s_j, \mathbf{x}_1 = x_i\right] = \Re(x_i) \langle g(t) \cos(\phi(s_j, t)), h_\ell(t) \rangle$$
$$- \Im(x_i) \langle g(t) \sin(\phi(s_j, t)), h_\ell(t) \rangle. \tag{27}$$

Moreover,

$$\langle g(t) \cos(\phi(s_j, t)), h_\ell(t) \rangle$$

$$= \int_0^T g^2(t) \sin(\phi(s_\ell, t)) \cos(\phi(s_j, t)) \, \mathrm{d}t \tag{28}$$

$$= \frac{1}{2} \int_0^T g^2(t) \sin(\phi(s_\ell + s_j, t)) \, \mathrm{d}t$$

$$+ \frac{1}{2} \int_0^T g^2(t) \sin(\phi(s_\ell - s_j, t)) \, \mathrm{d}t \tag{29}$$

$$= \Phi(s_\ell + s_j) + \Phi(s_\ell - s_j) \tag{30}$$

where we have set

$$\Phi(z) = \frac{1}{2} \int_0^T g^2(t) \sin(\phi(z, t)) \, \mathrm{d}t. \tag{31}$$

Similarly,

$$\langle g(t) \sin(\phi(s_j, t)), h_\ell(t) \rangle = \tilde{\Phi}(s_\ell - s_j) - \tilde{\Phi}(s_\ell + s_j) \tag{32}$$

where

$$\tilde{\Phi}(z) = \frac{1}{2} \int_0^T g^2(t) \cos(\phi(z, t)) \, \mathrm{d}t. \tag{33}$$

Therefore,

$$\mu_{ji\ell}^{\mathrm{R}} = \Re(x_i) \, \Phi(s_\ell + s_j) + \Re(x_i) \, \Phi(s_\ell - s_j)$$
$$+ \Im(x_i) \, \tilde{\Phi}(s_\ell + s_j) - \Im(x_i) \, \tilde{\Phi}(s_\ell - s_j). \tag{34}$$

With analogous calculations, we obtain

$$\tilde{\mu}_{ji\ell}^{\mathrm{R}} = \Re(x_i) \, \tilde{\Phi}(s_\ell + s_j) + \Re(x_i) \, \tilde{\Phi}(s_\ell - s_j)$$
$$- \Im(x_i) \, \Phi(s_\ell + s_j) + \Im(x_i) \, \Phi(s_\ell - s_j) \tag{35}$$

$$\mu_{ji\ell}^{\mathrm{I}} = -\Re(x_i) \, \tilde{\Phi}(s_\ell + s_j) + \Re(x_i) \, \tilde{\Phi}(s_\ell - s_j)$$
$$+ \Im(x_i) \, \Phi(s_\ell + s_j) + \Im(x_i) \, \Phi(s_\ell - s_j) \tag{36}$$

$$\tilde{\mu}_{ji\ell}^{\mathrm{I}} = \Re(x_i) \, \Phi(s_\ell + s_j) - \Re(x_i) \, \Phi(s_\ell - s_j)$$
$$+ \Im(x_i) \, \tilde{\Phi}(s_\ell + s_j) + \Im(x_i) \, \tilde{\Phi}(s_\ell - s_j). \tag{37}$$

Next, we calculate Σ, which is a $4|\mathcal{S}| \times 4|\mathcal{S}|$ matrix. Dividing Σ into 16 submatrices of size $|\mathcal{S}| \times |\mathcal{S}|$ and using [33, Prop. 25.15.2], we obtain

$$\Sigma = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} & 0 & 0 \\ \left(\Sigma^{12}\right)^T & \Sigma^{22} & 0 & 0 \\ 0 & 0 & \Sigma^{11} & \Sigma^{12} \\ 0 & 0 & \left(\Sigma^{12}\right)^T & \Sigma^{22} \end{bmatrix} \tag{38}$$

where the element $\Sigma_{k\ell}^{11}$ of the submatrix Σ^{11} is

$$\Sigma_{k\ell}^{11} = \frac{N_0}{2} \langle h_k(t), h_\ell(t) \rangle \tag{39}$$

$$= \frac{N_0}{2} \int_0^T g^2(t) \sin(\phi(s_k, t)) \sin(\phi(s_\ell, t)) \, \mathrm{d}t \tag{40}$$

$$= \frac{N_0}{4} \int_0^T g^2(t) \cos(\phi(s_k - s_\ell, t)) \, \mathrm{d}t$$
$$- \frac{N_0}{4} \int_0^T g^2(t) \cos(\phi(s_\ell + s_k, t)) \, \mathrm{d}t \tag{41}$$

$$= \frac{N_0}{2} \left[\tilde{\Phi}(s_k - s_\ell) - \tilde{\Phi}(s_k + s_\ell) \right] \tag{42}$$

for $k = 1, \ldots, |\mathcal{S}|$ and $\ell = 1, \ldots, |\mathcal{S}|$. Furthermore,

$$\Sigma_{k\ell}^{12} = \frac{N_0}{2} \langle h_k(t), \tilde{h}_\ell(t) \rangle = \frac{N_0}{2} \left[\Phi(s_k + s_\ell) + \Phi(s_k - s_\ell) \right] \tag{43}$$

$$\Sigma_{k\ell}^{22} = \frac{N_0}{2} \langle \tilde{h}_k(t), \tilde{h}_\ell(t) \rangle = \frac{N_0}{2} \left[\tilde{\Phi}(s_k + s_\ell) + \tilde{\Phi}(s_k - s_\ell) \right]. \tag{44}$$

Note that the real and the imaginary parts of $\mathbf{n}_1(t)$ are independent processes. This explains why half of the elements in (38) are zero.

E. MxM demodulation with MD detection (MxM-MD)

Next, we present a novel heuristic demodulation scheme, which is composed of three steps. First, the phase distortion of the received signal is estimated. This phase distortion is compensated for in the second step. Third, a MFS is applied to obtain the output of the demodulator. The first step is based on the following proposition, whose proof follows from [34, Ch. 4, Eq. (3)].

Proposition 1. Let $f(t)$ be a nonnegative continuous function on the interval $[a, b]$. Then

$$\max_{s \in \mathbb{R}} \left| \int_a^b f(t) e^{jsf(t)} \, \mathrm{d}t \right| = \int_a^b f(t) \, \mathrm{d}t \tag{45}$$

and $s = 0$ achieves the maximum.

Next, we use Proposition 1 to devise the first step of the demodulation. Assume that $\mathbf{x}_1 = x_1$ and $\mathbf{s} = s$. To estimate s, the receiver calculates

$$s_{\max} = \operatorname*{argmax}_{s' \in \mathcal{S}} \left| \int_0^T \mathbf{a}_1(L, t) \cdot g(t) e^{-j\eta_1 s' g^2(t)} \, \mathrm{d}t \right| \tag{46}$$

$$= \operatorname*{argmax}_{s' \in \mathcal{S}} \left| x_1 \int_0^T g^2(t) e^{j\eta_1 (s - s') g^2(t)} \, \mathrm{d}t + \mathbf{n} \right| \tag{47}$$

where $\mathbf{n} \sim \mathcal{CN}(0, N_0)$. If we ignore the noise in (47), it follows from Proposition 1 that $s_{\max} = s$. Therefore, $s_{\max}$ calculated in (46) provides an estimate of s in the presence of noise. Note that, similar to the SS decoder, the computation of $s_{\max}$ in (47) requires $4|\mathcal{S}|$ real-valued correlators.

In the next step, the phase distortion is compensated for by multiplying the received signal with $\exp\left(-j\eta_1 s_{\max} g^2(t)\right)$. Finally, the result is fed to the MFS demodulator. To summarize, the output of the MxM demodulator is

$$\mathbf{w} = \int_0^T \mathbf{a}_1(L, t) \cdot g(t) e^{-j\eta_1 s_{\max} g^2(t)} \, \mathrm{d}t. \tag{48}$$

We see from (8) that if the demodulator successfully compensates for the phase distortion, i.e., if $s_{\max} = s$, then the output of the MxM demodulator has a Gaussian distribution centered at x_1 with variance N_0. However, if s is not estimated correctly at the receiver, the output of the demodulator has a different mean. The MD detector determines $x_m \in \mathcal{X}$, based on the MxM output $\mathbf{w} = w$, such that

$$m = \arg \min_i |w - x_i|^2. \tag{49}$$

F. MxM demodulation with TS detection (MxM-TS)

To map the output of the MxM demodulator $\mathbf{w} = w$ in (48) to one of the constellation points, MxM-TS uses a simple two-stage detector. The two-stage detector first estimates the amplitude of the transmitted signal and then determines its phase. Specifically, let $\mathcal{R} = \{r_1, \ldots, r_{|\mathcal{R}|}\}$ be the set of all possible amplitudes of the transmitted symbol. The amplitude detector chooses $\hat{R} = r_i$, $1 \leq i \leq |\mathcal{R}|$, if $m_{i-1} \leq |w| \leq m_i$, where m_i is the ith detection threshold. To compute the thresholds m_i, we assume that given $\mathbf{x}_1 = x_1$, we have $\mathbf{w} \sim \mathcal{CN}(x_1, N_0)$. Therefore,

$$f_{|\mathbf{w}| \mid |\mathbf{x}_1|}(m_i \mid r_i) \approx \frac{2m_i}{N_0} \exp\left(-\frac{m_i^2 + r_i^2}{N_0} \right) I_0\left(\frac{2m_i r_i}{N_0} \right) \tag{50}$$

where $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind. Since the approximated conditional distribution in (50) is unimodal, m_i can be obtained based on the MAP rule by solving

$$\Pr(|\mathbf{x}_1| = r_i) f_{|\mathbf{w}| \mid |\mathbf{x}_1|}(m_i \mid r_i)$$
$$= \Pr(|\mathbf{x}_1| = r_{i+1}) f_{|\mathbf{w}| \mid |\mathbf{x}_1|}(m_i \mid r_{i+1}) \tag{51}$$

for $i = 1, \ldots, |\mathcal{R}| - 1$, (with the convention that $m_0 = 0$ and $m_{|\mathcal{R}|} = \infty$). After estimating the amplitude of the transmitted signal, the two-stage detector selects the constellation point with amplitude $\hat{R}$ that is closest to w.

G. Complexity

The MFS demodulator calculates the correlation between a real function $g(t)$ and the complex received signal, which can be implemented by two real-valued correlators (or, equivalently, filters). This number is $4|\mathcal{S}|$ for the

more sophisticated SS and MxM demodulators. The MAP detector in (24)–(25) involves calculating a quadratic form in a $4|\mathcal{S}|$–dimensional space, which makes it much more computationally demanding than the other detectors. The MAP detector in (14)–(16) involves calculating $|\mathcal{S}|$ exponential functions. Hence, it is more complex than the MD and TS detectors, which are only based on comparisons. Moreover, the MxM and the SS demodulators have larger bandwidths than the MFS demodulator. The bandwidth of the MxM demodulator is the maximum of the bandwidths of the signals $g(t)\exp(-j\eta s g^2(t))$ over all values of s; the bandwidth of the SS demodulator is the maximum of the bandwidths of the signals $h_\ell(t)$ and $\tilde{h}_\ell(t)$ over $\ell = 1,\ldots,|\mathcal{S}|$.

IV. ASYMPTOTIC SER ANALYSIS

In this section, we provide analytical evaluations of the asymptotic SER of the proposed receivers. Let the input random variable be $\mathbf{x}_1 = \sqrt{\mathcal{P}}\mathbf{x}'_1$, where $\mathbf{x}'_1$ takes values from a fixed alphabet set $\mathcal{X}' = \{x'_1, x'_2, ..., x'_{|\mathcal{X}'|}\}$, with some arbitrary probability distribution. Similarly, let $\mathbf{x}_2 = \sqrt{\mathcal{P}}\mathbf{x}'_2$. In order to make analytical calculations possible, we assume triangular pulse shaping, i.e.,

$$g(t) = c\left(\frac{T}{2} - \left|\frac{T}{2} - t\right|\right), \qquad c = \sqrt{\frac{12}{T^3}}. \qquad (52)$$

The following theorem presents our asymptotic SER results.

Theorem 1. Assuming triangular pulse shaping,
  i) the SER of the MFS-MAP receiver goes to $1 - \max_i(\pi_i)$ as $\mathcal{P} \to \infty$, where $\pi_i = \Pr(\mathbf{x}_1 = x_i)$.
  ii) the SER of the MxM-MD and MxM-TS receivers goes to zero as $\mathcal{P} \to \infty$.

Proof: Substituting (8) into (9), we can write the output of the MFS demodulator as

$$\mathbf{v} = \int_0^T \mathbf{x}_1 g^2(t) \exp\left(j\eta_1 \left(|\mathbf{x}_1|^2 + 2|\mathbf{x}_2|^2\right) g^2(t)\right) \mathrm{d}t + \mathbf{n} \,(53)$$

$$= 2c^2 \mathbf{x}'_1 \int_0^{T/2} \sqrt{\mathcal{P}} t^2 \exp\left(j\eta_1 c^2 \mathcal{P}\left(|\mathbf{x}'_1|^2 + 2|\mathbf{x}'_2|^2\right) t^2\right) \mathrm{d}t$$
$$+ \mathbf{n}. \qquad (54)$$

where $\mathbf{n} \sim \mathcal{CN}(0, N_0)$. Here, (54) follows from (52) and the definitions of $\mathbf{x}'_1$ and $\mathbf{x}'_2$. We first assume $\mathbf{x}'_1 \neq 0$. It can be shown by standard algebraic calculations that the integral in (54) is $\mathcal{O}\left(1/\sqrt{\mathcal{P}}\right)$. Therefore, as $\mathcal{P} \to \infty$ the first term in (54) goes to zero. Furthermore, this term is zero if $\mathbf{x}'_1 = 0$. Since $\mathbf{n}$ is independent of the transmitted signal, the MAP detector selects, in the limit $\mathcal{P} \to \infty$, the symbol with largest a priory probability regardless of received signal, resulting in a SER of $1 - \max_i(\pi_i)$.

Next, we prove the second part of the theorem. Focusing on (47), one can show with similar calculations as above that for every $s \neq s'$, the integral $x_1 \int g^2(t)e^{j\eta_1(s-s')g^2(t)} \mathrm{d}t$ goes to zero as $\mathcal{P} \to \infty$. Moreover, for $s = s'$, the integral equals x_1. Therefore, assuming

Parameter	Symbol	Value
Span length	L_{span}	150 km
Attenuation	α	0.25 dB/km
Nonlinearity	$\gamma_1 = \gamma_2$	1.27 $(\mathrm{Wkm})^{-1}$
Symbol rate	$1/T$	10 Gbaud
Optical photon energy	$h\nu$	$1.28 \cdot 10^{-19}$ J
Amplifier noise figure	F	6 dB
Number of spans	n_{span}	1

$x_1 \neq 0$, we conclude that, in the limit $\mathcal{P} \to \infty$, we have $s_{\max} = s$ with probability one. Under the assumption that $s_{\max} = s$, it follows from (48) that $\mathbf{w} \sim \mathcal{CN}\left(\sqrt{\mathcal{P}}x'_1, N_0\right)$, where $\mathbf{w}$ is the outcome of the MxM demodulator. Therefore, in the limit $\mathcal{P} \to \infty$, both MD and TS detectors will correctly detect the symbol x_1 with probability one. If $x_1 = 0$, then $\mathbf{w}$ does not depend on $\mathcal{P}$, and therefore in the limit $\mathcal{P} \to \infty$, the symbol 0 will be correctly detected by both MD and TS with probability one. ∎

Note that the first result in Theorem 1 implies that the asymptotic SER of the MFS-MD and MFS-PR is lower-bounded by $1 - \max_i(\pi_i)$; the second result implies that the SER of SS-MAP goes to zero as $\mathcal{P} \to \infty$.

V. NUMERICAL EXAMPLES

This section presents numerical SER evaluations for three single-span channels. The simplified model (3)–(4) is studied in Section V-A and two NLS channels are analyzed via split-step Fourier simulations in Section V-B.

A. Transmission Over the Simplified Channel (3)–(4)

In this section, we evaluate the performance of the six receivers presented in the previous section, by conducting Monte Carlo simulations on the channel model (3)–(4). We consider the transmission of 16-ary quadrature amplitude modulation (QAM) data symbols from each of the two transmitters. The input power $\mathcal{P} = E_s/T$, where $E_s = \mathrm{E}[|\mathbf{x}_1|^2]$, is assumed to be the same for both channels. For these choices we have that $|\mathcal{S}| = 7$ and

$$\mathcal{S} = \{0.6E_s, \ 1.4E_s, \ 2.2E_s, \ 3E_s, \ 3.8E_s, \ 4.6E_s, \ 5.4E_s\}. \qquad (55)$$

The simulation parameters can be found in Table II. The nonlinear coefficient can be calculated from (5)–(6) as $\eta_1 = \eta_2 = 22.1$ W^{-1}. Also, using (7), one obtains $N_0 = 1.43 \cdot 10^{-15}$ W/Hz. We use 100 samples per symbol and set $g(t)$ to a truncated Gaussian pulse with a full width at half maximum of $T/2$. A uniform input distribution is assumed for both transmitted signals, i.e., $\pi_i = 1/16$. Consequently, the conditional probabilities $\tilde{\pi}_{ji}$ can be calculated as in Table III.

Fig. 1 shows scatter plots of the MFS and MxM demodulator outputs for three levels of input power. Note that since the output of the SS demodulator (19)–(22) lies in a
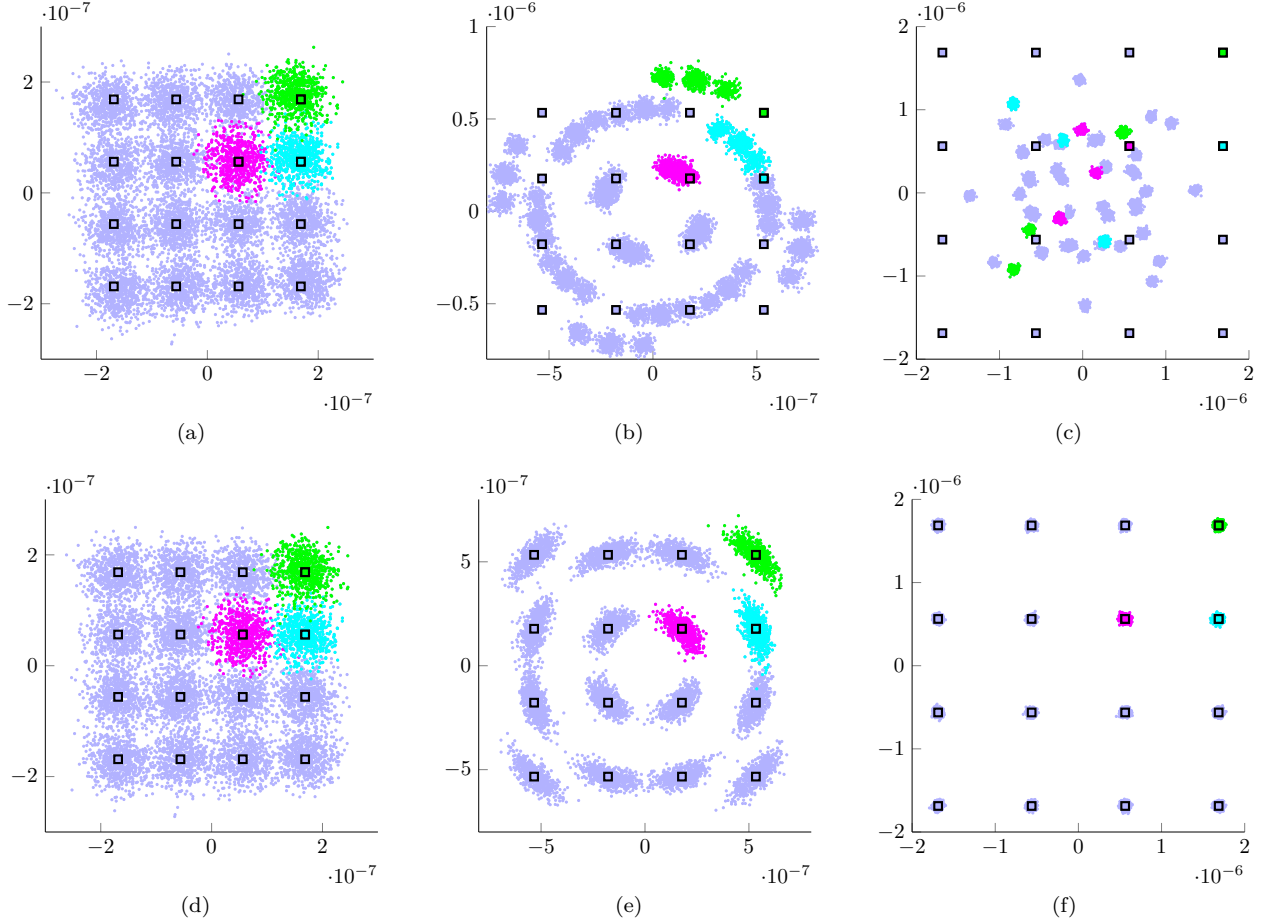
Fig. 1. Scatter plots of the output of two demodulation schemes, matched filtering and sampling (MFS) and maximum matching (MxM), for 16-QAM with three input powers $\mathcal{P} = -5$ dBm: (a) and (d), $\mathcal{P} = 5$ dBm: (b) and (e), and $\mathcal{P} = 15$ dBm: (c) and (f). Different colors are used to identify demodulator outputs corresponding to three given input symbols.

TABLE III
$\tilde{\pi}_{ji} = \Pr(\mathbf{s} = s_j \mid |\mathbf{x}| = |x_i|)$ for 16-QAM transmission with uniform distribution.

$|x_i|$ \ s_j	$0.6E_s$	$1.4E_s$	$2.2E_s$	$3E_s$	$3.8E_s$	$4.6E_s$	$5.4E_s$
$\sqrt{0.2E_s}$	0.25	0	0.5	0	0.25	0	0
$\sqrt{E_s}$	0	0.25	0	0.5	0	0.25	0
$\sqrt{1.8E_s}$	0	0	0.25	0	0.5	0	0.25

vector space with dimension $4|\mathcal{S}| = 28$, it is not possible to draw its scattering pattern. At $\mathcal{P} = -5$ dBm, it can be seen from Fig. 1(a) that the output of the MFS demodulation follows approximately a Gaussian distribution. However, the clouds are not centered at the constellation points. Rather, they are rotated by an amount proportional to the amplitude square of the constellation points. This rotation is caused by the SPM distortion. In Fig. 1(d), the output of the MxM demodulator at $\mathcal{P} = -5$ dBm is shown. It can be seen that with this demodulator, the effect of SPM is mitigated. Indeed, the clouds are now centered at the constellation points.

Fig. 1(b) illustrates the MFS demodulator's output at $\mathcal{P} = 5$ dBm. One can observe that the effect of the

nonlinear distortion becomes more significant compared to the case $\mathcal{P} = -5$ dBm. Each constellation point is scattered to three different clouds, each one corresponding to the three possible values of the XPM distortion (the three values of $|\mathbf{x}_2|$). Also, the centers of the clouds are further rotated away from the constellation points because of the SPM. As shown in Fig. 1(e), the output of MxM is also dispersed to three clouds per symbol. However, unlike MFS, these clouds are centered at the constellation points.

One can observe from Fig. 1(c) that when $\mathcal{P} = 15$ dBm both the phase and the amplitude of the MFS output are distorted. The power loss, which is evident in Fig. 1(c), can be explained as follows. At high powers, the phase of the integrand in (9) changes quickly during one time slot. This rapid phase change scales down the integral's result in (9), which is the output of the MFS demodulator. Alternatively, the power loss can be explained in the frequency domain. At high powers, the nonlinear distortions substantially broaden the signal's spectrum. However, MFS uses a filter matched to the transmitted pulse shape, which has the same bandwidth as the transmitted signal. Therefore, the signal's out-of-band energy is excluded. It can be seen from Fig. 1(f) that the output of the MxM demodulator is centered at each constellation point, i.e., there is no
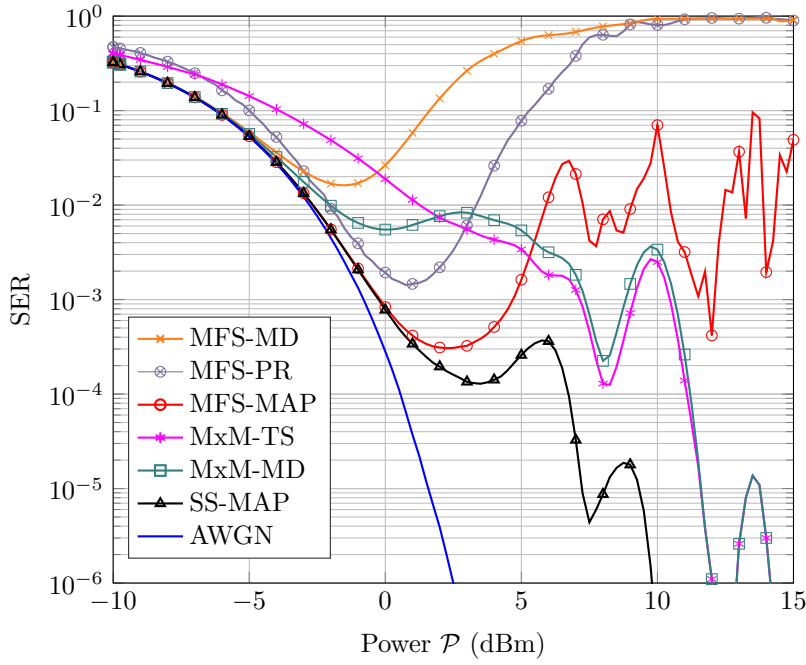
Fig. 2. The SER of the six receivers introduced in Section III is illustrated by conducting Monte-Carlo simulations on the channel model (3)–(4). The SER of an AWGN channel with the same noise variance is also plotted for comparison.

power loss or phase distortion. Fig. 1(f) indicates that the nonlinear distortion is effectively compensated for by the MxM demodulator.

Fig. 2 depicts the SER for the six receivers introduced in Section III. Moreover, the SER for the AWGN channel, obtained by setting $\eta_1 = 0$ in (8), is plotted for comparison. In the following, we discuss the results in Fig. 2 for each demodulation scheme.

MFS demodulator: In our analysis, this demodulator is combined with two detectors, namely, MD and MAP. It is well known that for the AWGN channel and a uniform input distribution, these two detectors coincide. On the contrary, it can be observed in Fig. 2 that for the nonlinear channel considered here, a substantial gap exists between the performance of these two detectors. The SER for the MFS-MD receiver follows first the SER of the AWGN channel, reaches a minimum point of $1.6 \cdot 10^{-2}$, and then increases to approximately one at high power levels. The increase in the SER in the high-power regime can be explained by looking at Figs. 1(a)–(c). The output of the MFS demodulator is not centered at the constellation points. Therefore, the MD decoder fails to provide a sound estimate of the transmitted symbols. Comparing MFS-PR with MFS-MD, it can be seen that a considerable improvement is obtained by performing phase recovery. The minimum SER for MFS-PR is $1.4 \cdot 10^{-3}$.

By changing the detection scheme from MD to MAP, a substantial performance gain can be obtained. The MFS-MAP receiver yields a SER of $3.1 \cdot 10^{-4}$ at $\mathcal{P} = 2$ dBm, which is more than 50 times smaller than the minimum SER that can be obtained with the MFS-MD. The MAP detector can identify the transmitted symbols as long as the output of the MFS consists of well-separated clouds.

However, as shown in Figs. 1(a)–(c), because of the nonlinearity, the clouds move in the constellation plane as the power level changes and can overlap. Therefore, based on the position of the clouds, increasing the input power can enhance or deteriorate the performance, which causes the somewhat irregular behavior of the SER for the MFS-MAP receiver in Fig. 2.

MxM demodulator: Two detector schemes, namely, TS and MD, are combined with the MxM demodulator. It can be seen in Fig. 2 that at power levels lower than 2 dBm, MD outperforms TS; when 2 dBm $\leq \mathcal{P} \leq 11$ dBm, TS yields a smaller SER than MD; and at power levels larger than 11 dBm, both detectors perform equally. The reason is as follows. In the low-power regime, the nonlinearity is weak and the output of the MxM demodulator has approximately a Gaussian distribution centered at the transmitted signal (see Fig. 1(d)). Therefore, MD detection is close to optimal at low powers. In the moderate-power regime, the output of the MxM demodulator experiences a phase distortion caused by SPM and XPM (see Fig. 1(e)). In the presence of phase distortion, TS outperforms MD, as previously reported in the literature (see [25], for example). Next, we explain why the MxM-MD and the MxM-TS receivers yield the same SER at high powers. The MxM demodulator first tries to cancel the nonlinear distortion. If it succeeds, the output of the demodulator follows a Gaussian distribution centered at the transmitted symbol. Otherwise, the outcome of MxM gets distorted by the nonlinearity. In the first case, both the MD and TS detectors are able to detect the transmitted symbol almost without error. In the second case, both detectors make most likely an error because the phase and the amplitude of the demodulator output
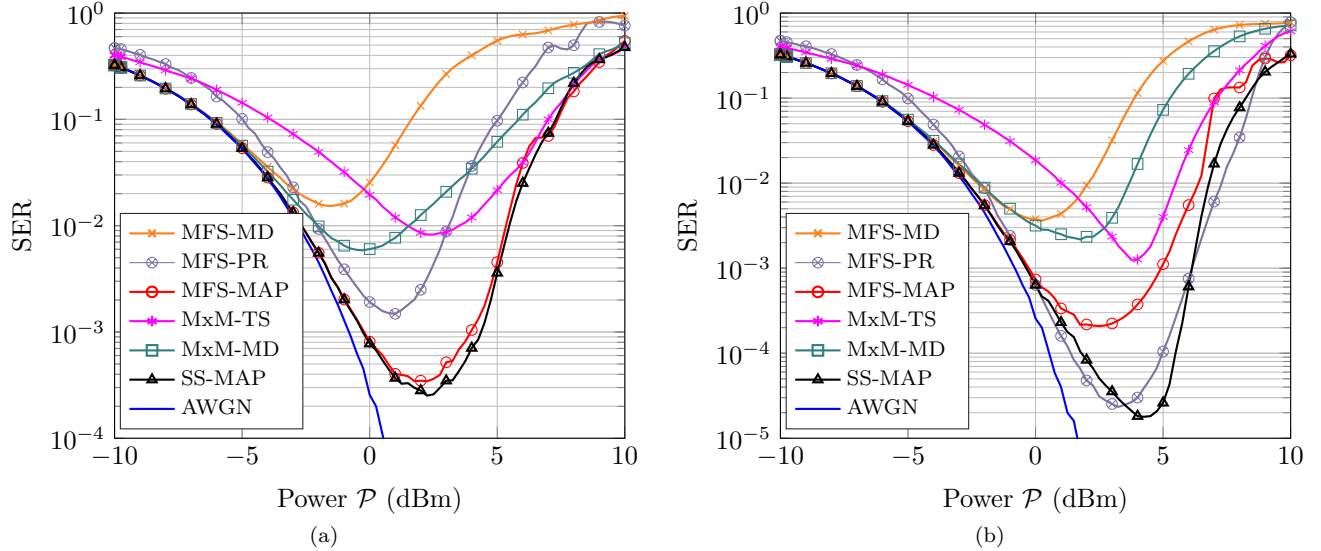
Fig. 3. The SER of the six mismatched receivers introduced in Section III is plotted for a single-span transmission with (a) low-dispersive fiber and (b) standard fiber. The SER of an AWGN channel with the same noise variance is also plotted for comparison.

are severely distorted at high powers. This also causes the nonmonotonic behaviour of the SER as a function of the power.

SS-MAP receiver: SS-MAP is the optimal receiver for the channel under study (although it has a high complexity) and its SER can serve as a benchmark to compare the performance of other low-complexity receivers. One can see that the SER of the MFS-MAP follows that of the optimal receiver closely up until $\mathcal{P} = 1$ dBm. However, unlike the MFS-MAP, the SER of the SS-MAP and of both the MxM receivers vanishes at high power levels. We see from Fig. 2 that the effect of the nonlinearity cannot be completely mitigated even by using optimal demodulation and detection schemes, as there exists a considerable gap between the SER of the SS-MAP receiver and SER achievable over an AWGN channel.

B. Transmission Over Two Single-Span NLS Channels

In this section, we evaluate the performance of the receivers introduced in Section III for two realistic single-span fiber-optical systems, one with a low-dispersion SMF and the other with a standard SMF. We use the MFS-PR receiver as a benchmark. We note that the SS demodulator and the MAP detector no longer represent the optimal demodulation and detection schemes, as they have been designed for the simplified channel model (3)–(4) and are mismatched to the channel under study in this section.

The signals $\mathbf{a}_1(0,t)$ and $\mathbf{a}_2(0,t)$ are passed through a brick-wall filter with bandwidth $\Delta f/2$, where Δf is the channel spacing parameter in hertz. The baseband input signal, $\mathbf{a}(0,t)$, is generated according to

$$\mathbf{a}(0,t) = \mathbf{a}_1(0,t)e^{-j\pi t\Delta f} + \mathbf{a}_2(0,t)e^{j\pi t\Delta f} \qquad (56)$$

The input signal $\mathbf{a}(0,t)$ is transmitted through the fiber-optical channel governed by the NLS equation

$$\frac{\partial \mathbf{a}}{\partial z} + \frac{j\beta_2}{2}\frac{\partial \mathbf{a}}{\partial t^2} + \frac{\alpha}{2}\mathbf{a} = j\gamma|\mathbf{a}|^2 \qquad (57)$$

where β_2, γ, and α are dispersion, nonlinearity, and attenuation coefficients, respectively. The fiber loss is compensated completely by an optical amplifier. The dispersion is compensated at each receiver digitally.

We consider two fiber-optical systems with different dispersion parameters. The first system deploys a quadruply clad fiber [31, Ch. 1] with $\beta_2 = -1.27$ ps^2/km and the second system uses a standard SMF with $\beta_2 = -21.7$ ps^2/km. The channel spacing parameter is $\Delta f = 40$ GHz. The values of the other parameters can be found in Table II. The solution of (57) is approximated by the split-step Fourier method [31, Ch. 2.4.1]; 100 samples are taken from each symbol to discretize the input signal. Pulse shaping is the same as in Section V.

Fig. 3(a) illustrates the performance of the six receivers for a low-dispersive fiber. It can be seen that because of the nonlinearity–dispersion interplay, the SER of all the receivers increases after reaching a global minimum. By using the SS-MAP and the MFS-MAP receivers, considerable performance gains can be achieved compared to MFS-PR. One can see that the MxM-MD and MxM-TS perform worse than MFS-PR but better than the MFS-MD receiver. Fig. 3(b) presents the SER for a standard SMF. The dispersion is high and the MFS-PS and SS-MAP perform better than the other receivers.

VI. CONCLUSION AND DISCUSSION

Six receivers were studied for a two-user simplified WDM channel and a novel demodulator, referred to as MxM, was proposed. Our results indicate that the MFS-MD receiver, which is optimal for the AWGN channel,

performs very poorly in the presence of optical nonlinear distortion. However, when the output of the MFS is fed to a MAP detector, one can achieve performance close to the optimal receiver at low powers. In the high-power regime, the SER goes to zero with power for the optimal receiver as well as for the receivers based on the MxM demodulator. On the contrary, for receivers based on the MFS demodulator, the SER does not vanish.

In coherent optical transmissions the signal spectrum broadens at high transmit power levels, because of the nonlinearity. The information embedded in the out-of-band frequencies is however ignored by the MFS demodulator. Our results indicate that ignoring this information loss deteriorates performance substantially at high powers. Moreover, by proposing the MxM demodulator, we showed that a vanishing SER can be obtained by a heuristic receiver that is simpler than the optimal one.

When evaluated over a more realistic single-span fiber-optical channel, modeled by the NLS equation, the performance of all receivers declines in the high-power regime. In the low-dispersion case two of the receivers analyzed in this paper, namely MFS-MAP and SS-MAP outperform the conventional MFS-PR receiver. Since the receivers in this paper were designed based on a simplified memoryless model, further improvement is expected by devising receivers that take into account both dispersion and nonlinearity. It seems that developing the optimal receiver in the presence of dispersion is a formidable task and heuristic methods should be considered. A straightforward approach may be optimizing the performance of the proposed receivers over different values of η_1. Since dispersion mitigates the effects of nonlinearity, the optimal η_1 may be smaller than the right-hand side of (5).

Finally, we note that equalization and phase recovery are essential parts of today's optical receivers. While the performance of the introduced receivers may be influenced by these two steps, we have not investigated the proper coupling of the equalization and the phase-recovery processes with the demodulation and detection steps. This is an interesting topic for future studies.

References

[1] K. Keykhosravi and E. Agrell, "A novel demodulation scheme for a memoryless optical interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, June 2017, pp. 66–70.

[2] T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita, "Ultimate low-loss single-mode fibre at 1.55 μm," *Electron. Lett.*, vol. 15, no. 4, pp. 106–108, Feb. 1979.

[3] R. J. Mears, L. Reekie, I. Jauncey, and D. N. Payne, "Low-noise erbium-doped fibre amplifier operating at 1.54 μm," *Electron. Lett.*, vol. 23, no. 19, pp. 1026–1028, Sep. 1987.

[4] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[5] F. Forghieri, R. Tkach, A. R. Chraplyvy, and D. Marcuse, "Reduction of four-wave mixing crosstalk in WDM systems using unequally spaced channels," *IEEE Photon. Technol. Lett.*, vol. 6, no. 6, pp. 754–756, Jun. 1994.

[6] N. V. Irukulapati, "Towards the limits of nonlinearity compensation for fiber-optic channels," Ph.D. dissertation, Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden, 2016.

[7] P. D. Lax, "Integrals of nonlinear equations of evolution and solitary waves," *Commun. Pure Appl. Math.*, vol. 21, no. 5, pp. 467–490, Sep. 1968.

[8] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear Fourier transform, part III: Spectrum modulation," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4346–4369, Jul. 2014.

[9] R. A. Fisher, B. Suydam, and D. Yevick, "Optical phase conjugation for time-domain undoing of dispersive self-phase-modulation effects," *Opt. Lett.*, vol. 8, no. 12, pp. 611–613, Dec. 1983.

[10] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, Oct. 2008.

[11] M. Secondini and E. Forestieri, "On XPM mitigation in WDM fiber-optic systems," *IEEE Photon. Technol. Lett.*, vol. 26, no. 22, pp. 2252–2255, Nov. 2014.

[12] V. Arlunno, X. Zhang, K. J. Larsen, D. Zibar, and I. T. Monroy, "Digital non-linear equalization for flexible capacity ultradense wdm channels for metro core networking," *Optics express*, vol. 19, no. 26, pp. B270–B276, Dec. 2011.

[13] D. Marsella, M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *J. Lightw. Technol.*, vol. 32, no. 5, pp. 908–916, Mar. 2014.

[14] Y. Cai, D. G. Foursa, C. R. Davidson, J.-X. Cai, O. Sinkin, M. Nissov, and A. Pilipetskii, "Experimental demonstration of coherent MAP detection for nonlinearity mitigation in long-haul transmissions," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, Mar. 2010, paper OTuE1.

[15] G. P. Agrawal, *Fiber-Optic Communication Systems*, 3rd ed. New York, NY: John Wiley & Sons, 2002.

[16] A. S. Tan, H. Wymeersch, P. Johannisson, E. Agrell, P. A. Andrekson, and M. Karlsson, "An ML-based detector for optical communication in the presence of nonlinear phase noise," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011.

[17] A. P. T. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *J. Lightw. Technol.*, vol. 25, no. 10, pp. 3008–3016, Oct. 2007.

[18] K. Keykhosravi, G. Durisi, and E. Agrell, "A tighter upper bound on the capacity of the nondispersive optical fiber channel," in *Proc. European Conference on Optical Communication (ECOC)*, Gothenburg, Sweden, Sep. 2017. [Online]. Available: http://publications.lib.chalmers.se/publication/250961

[19] M. I. Yousefi and F. R. Kschischang, "On the per-sample capacity of nondispersive optical fibers," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.

[20] L. Beygi, E. Agrell, M. Karlsson, and P. Johannisson, "Signal statistics in fiber-optical channels with polarization multiplexing and self-phase modulation," *J. Lightw. Technol.*, vol. 29, no. 16, pp. 2379–2386, Aug. 2011.

[21] E. Agrell, G. Durisi, and P. Johannisson, "Information-theory-friendly models for fiber-optic channels: A primer," in *IEEE Inf. Theory Workshop (ITW)*, Apr. 2015.

[22] K. Keykhosravi, G. Durisi, and E. Agrell, "Bounds on the capacity of memoryless simplified optical channel models," *arXiv:1708.03102*, Aug. 2017.

[23] H. Ghozlan and G. Kramer, "Interference focusing for mitigating cross-phase modulation in a simplified optical fiber model," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, Jun. 2010, pp. 2033–2037.

[24] ——, "Interference focusing for simplified optical fiber models with dispersion," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, July–Aug. 2011, pp. 376–379.

[25] C. Häger, A. Graell i Amat, A. Alvarado, and E. Agrell, "Design of APSK constellations for coherent optical channels with nonlinear phase noise," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3362–3373, Aug. 2013.

[26] G. Kramer, "Autocorrelation function for dispersion-free fiber channels with distributed amplification," *arXiv preprint: 1705.00454*, 2017.

[27] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.

[28] M. Secondini and E. Forestieri, "Analytical fiber-optic channel model in the presence of cross-phase modulation," *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016–2019, Nov. 2012.

[29] L. Beygi, E. Agrell, and M. Karlsson, "Optimization of 16-point ring constellations in the presence of nonlinear phase noise," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Los Angeles, CA, 2011, paper OThO4.

[30] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations," *J. Lightw. Technol.*, vol. 27, no. 8, pp. 989–999, Apr. 2009.

[31] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. New York, NY: Academic Press, 2007.

[32] ——, *Nonlinear Fiber Optics*, 3rd ed. San Diego, CA: Academic Press, 2001.

[33] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[34] L. V. Ahlfors, *Complex Analysis, an Introduction to the Theory of Analytic Functions of One Complex Variable*, 3rd ed. New York, NY: McGraw-Hill Book Co., 1979.

Erik Agrell (M'99–SM'02–F'18) received the Ph.D. degree in information theory in 1997 from Chalmers University of Technology, Sweden.

From 1997 to 1999, he was a Postdoctoral Researcher with the University of California, San Diego and the University of Illinois at Urbana-Champaign. In 1999, he joined the faculty of Chalmers University of Technology, where he is a Professor in Communication Systems since 2009. In 2010, he cofounded the Fiber-Optic Communications Research Center (FORCE) at Chalmers, where he leads the Electrical Engineering research area. He was a Visiting Professor at University College London in 2014–2017. His research interests belong to the fields of information theory, coding theory, and digital communications, and his favorite applications are found in optical communications.

Prof. Agrell served as Publications Editor for the IEEE Transactions on Information Theory from 1999 to 2002 and as Associate Editor for the IEEE Transactions on Communications from 2012 to 2015. He is a recipient of the 1990 John Ericsson Medal, the 2009 ITW Best Poster Award, the 2011 GlobeCom Best Paper Award, the 2013 CTW Best Poster Award, the 2013 Chalmers Supervisor of the Year Award, and the 2018 JLT Best Paper Award.

Kamran Keykhosravi (S'16) received the B.Sc. and M.Sc. degree both in electrical engineering in 2012 and 2014, respectively, from Sharif University of Technology, Iran. In 2015, he joined the Communication systems and information theory group, Chalmers University of Technology as a PhD student. His main research interests include applications of information theory in fiber optical communications.

Giuseppe Durisi (S'02–M'06–SM'12) received the Laurea degree summa cum laude and the Doctor degree both from Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2006 to 2010 he was a postdoctoral researcher at ETH Zurich, Zurich, Switzerland. In 2010, he joined Chalmers University of Technology, Gothenburg, Sweden, where he is now professor and co-director of Chalmers information and communication technology Area of Advance. Dr. Durisi is a senior member of the IEEE. He is the recipient of the 2013 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East, and Africa Region, and is co-author of a paper that won a "student paper award" at the 2012 International Symposium on Information Theory, and of a paper that won the 2013 IEEE Sweden VT-COM-IT joint chapter best student conference paper award. In 2015, he joined the editorial board of the IEEE Transactions on Communications as associate editor. From 2011 to 2014, he served as publications editor for the IEEE Transactions on Information Theory. His research interests are in the areas of communication and information theory.

Morteza Tavana (S'17) Morteza Tavana received the B. Sc. degree in electrical engineering from Amirkabir University of Technology, Tehran, Iran, in 2012 and the M. Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2015. In 2016, he joined the Communication systems group, Chalmers University of Technology as a PhD student. His main research interests include digital communication, signal processing, detection and estimation theory and their applications on fiber optical communications.