Chalmers Publication Library

(article starts on next page)

Impact of 4D Channel Distribution on the Achievable Rates in Coherent Optical Communication Experiments

Tobias A. Eriksson, Tobias Fehenberger, Peter A. Andrekson, Magnus Karlsson, Norbert Hanik, and Erik Agrell

Abstract—We experimentally investigate mutual information and generalized mutual information for coherent optical transmission systems. The impact of the assumed channel distribution on the achievable rate is investigated for distributions in up to four dimensions. Single channel and wavelength division multiplexing (WDM) transmission over transmission links with and without inline dispersion compensation are studied. We show that for conventional WDM systems without inline dispersion compensation, a circularly symmetric complex Gaussian distribution is a good approximation of the channel. For other channels, such as with inline dispersion compensation, this is no longer true and gains in the achievable information rate are obtained by considering more sophisticated four-dimensional (4D) distributions. We also show that for nonlinear channels, gains in the achievable information rate can also be achieved by estimating the mean values of the received constellation in four dimensions. The highest gain for such channels is seen for a 4D correlated Gaussian distribution.

Index Terms—Channel models, Digital communication, Fiber nonlinear optics, Mutual Information, Optical fiber communication.

I. INTRODUCTION

COHERENT optical communication systems have largely been enabled by the use of digital signal processing (DSP) [1] which is used to mitigate signal distortions such as laser phase drifts and polarization drifts. This has eased the use of higher order modulation formats that make use of both the phase and the amplitude of the optical field. The driving force to increase the modulation order is the ever-increasing demand for increased data rates in today's communication systems [2], as a higher order modulation format enables transmission with a higher spectral efficiency (SE).

One of the key technologies in present communication systems is forward error correction (FEC) coding. The use of FEC can increase the sensitivity of a communication system significantly at the cost of lower spectral efficiency, due to

Tobias A. Eriksson, Peter A. Andrekson, and Magnus Karlsson are with the Photonics Laboratory, Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden (e-mail: tobias.eriksson@chalmers.se)

Tobias Fehenberger and Norbert Hanik are with the Institute for Communications Engineering, Technical University of Munich, 80333 Munich, Germany

Erik Agrell is with the Communication Systems Group, Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden.

the overhead from the code, and increased complexity in the encoding and decoding circuitry. Without the use of FEC, many higher-order modulation formats such as polarization-multiplexed 16-ary quadrature amplitude modulation (PM-16QAM) cannot be transmitted over any significant distance before the detected data can no longer be considered "error-free" (typically defined as bit error rate (BER) $< 10^{-15}$) [3]. However, using advanced FEC, PM-16QAM can be transmitted over transoceanic distances [4]. The use of advanced FEC in combination with PM quadrature phase shift keying (PM-QPSK) has enabled record SE-transmission distance products [5], [6].

Up until recent years, the FEC codes used in fiber-optic transmission systems were typically Reed-Solomon (RS) [7] or Bose-Chaudhuri-Hocquenghem (BCH) codes [8], decoded using hard-decision (HD) algorithms. This means that the received constellation is demapped into bits before information is passed to the decoder. In recent years, the FEC codes considered for fiber-optic systems are based on soft-decision (SD) decoding, which means that the soft information from the received constellation after the receiver DSP is sent to the decoder. Examples of such codes are low-density parity check codes (LDPC) [9], polar codes [10], and turbo codes [11]. These types of codes can achieve significantly higher sensitivity compared to the HD codes at the cost of higher complexity. It should be noted that in some cases, concatenated codes with an inner SD code and an outer HD code are considered, which is practical if the inner code has an error-floor above the error-free BER limit.

For HD coding schemes, if the channel seen by the FEC can be modeled as a binary symmetric channel, the achievable rate after decoding can be fully determined by the pre-FEC BER [12]. In other words, estimating the BER from the received constellation gives a good estimate of the post-FEC BER and is indeed what is done in most fiber-optic transmission experiments. The main reason that the FEC decoder is not implemented in experiments is that the number of samples needed for good statistics at a post-FEC BER of 10^{-15} is infeasible with offline processing. For SD decoders on the other hand, there exists no such relation between the pre-FEC BER and the achievable rate. This can be explained by the fact that an SD decoder does not work on bits as input. For optical communication systems with SD FEC, it has been shown that mutual information (MI) is a more reliable measure than the pre-FEC BER [13]. In [14], it is shown that for an LDPC code and a turbo-product code (TPC) with bit-wise decoders, the
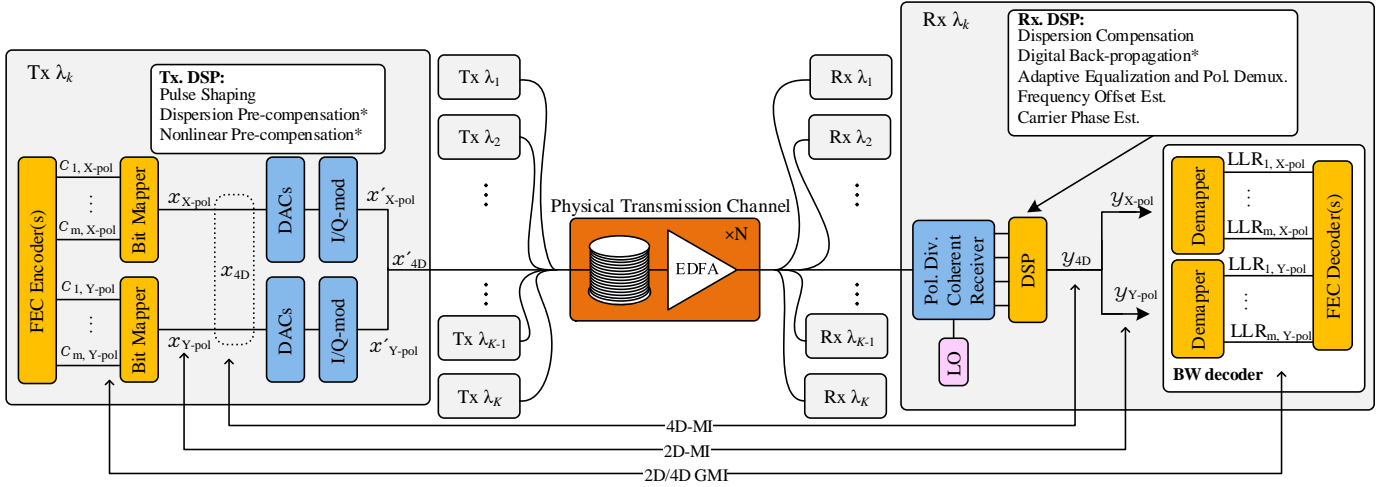
Fig. 1: Schematic of a typical fiber optical communication system. Items listed under Tx. DSP and Rx. DSP functions that are marked with * are not implemented in this work.

generalized mutual information (GMI) gives a good estimate of the post-FEC BER independently of the modulation format and optical launch power for a link without inline dispersion compensation.

Most of the previous work on mutual information as a figure of merit in fiber-optic communication systems has been carried out theoretically or with simulations [14]–[20], or in experiments where the output statistics of the channel is considered as memoryless with additive Gaussian noise statistics with the same variance in all dimensions [13], [21]–[23].

However, the fiber-optic channel is inherently nonlinear, which when approximated as memoryless white Gaussian noise means that information is lost. This has been discussed in several publications [16], [17], [19], [20], [24]. In [25], we study the impact of different four-dimensional (4D) assumption of the memoryless channel distribution on the achievable information rate. We show that for single-channel transmission of 14 Gbaud PM-16QAM in a transmission system with inline dispersion compensation with high launch powers, significant gains in the achievable information rate can be achieved by employing receiver-side models that assume 4D channel distributions.

In this work we experimentally investigate different estimates of the achievable rate using both MI and GMI. We compare different models of the channel distribution and its impact on the achievable rate for different transmission scenarios. We show that for most realistic scenarios, circularly symmetric Gaussian noise statistics is a good assumption. We also show that for some specific links, a higher achievable rate can be achieved by using more sophisticated channel models, based on 4D distributions, in the decoder.

II. ACHIEVABLE INFORMATION RATES

Fig. 1 shows a schematic of a typical long-haul point-to-point fiber optical communication system for which different estimates of the achievable information rate will be derived in this section.

A. Mutual Information

We follow the approach of [17], [18], [26] to evaluate a lower bound estimate on the symbolwise mutual information (MI) from real d-dimensional symbols that are obtained after transmission and DSP. Let the channel input $\boldsymbol{X}$ be a d-dimensional random variable that is drawn from a constellation $\mathcal{X} = \{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$ with cardinality $|\mathcal{X}| = M = 2^m$ with identical probability. The channel output $\boldsymbol{Y}$ denotes a random variable that is dependent on $\boldsymbol{X}$ and the channel. The MI between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given as

$$I(\boldsymbol{X}; \boldsymbol{Y}) = \mathrm{E}_{\boldsymbol{XY}} \left[\log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{p_{\boldsymbol{Y}}(\boldsymbol{Y})} \right], \quad (1)$$

where $p_{\boldsymbol{Y}|\boldsymbol{X}}$ is the conditional memoryless channel transition rldistribution and $p_{\boldsymbol{Y}}$ the channel output density. The MI gives the highest rate for a specific channel and input distribution at which reliable communication is possible and cannot exceed m bits per channel use. In other words, below the MI there exist codes with the possibility of a post-FEC BER that approaches 0. As indicated in Fig. 1, the MI is calculated on a symbol level, i.e., the bit-to-symbol mapping does not influence the MI.

Note that (1) gives the MI for a memoryless channel. In reality, a fiber-optic channel exhibits memory, which makes $I(\boldsymbol{X}; \boldsymbol{Y})$ a lower bound on the MI calculated with the channel input and output as sequences [27, Sec. III-F] [28]. We also note that considering a finite memory due to nonlinear effects of the fiber-optic channel can have a large effect on the achievable information rate [19]. Since most of the linear memory introduced in the optical fiber channel, such as dispersion and inter-symbol interference, is compensated by the DSP, we restrict our analysis to the memoryless MI for the remainder of this work. The experimental results in Section IV nevertheless apply to the true fiber-optic channel with residual memory after DSP.

Using the weak law of large numbers, we can estimate $I(\boldsymbol{X}; \boldsymbol{Y})$ of (1) via Monte-Carlo integration from N input-

output pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ [26] as

$$\frac{1}{N} \sum_{i=1}^{N} \log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{x}_i)}{p_{\boldsymbol{Y}}(\boldsymbol{y}_i)} \xrightarrow{p} I(\boldsymbol{X}; \boldsymbol{Y}), \qquad (2)$$

where $\xrightarrow{p}$ denotes convergence in probability. Eq. (2) implies that if there is an analytical description of the channel over which we have transmitted N symbols, an estimate of the MI is obtained whose accuracy increases with N.

In fiber optical communications, however, the channel transition distribution $p_{\boldsymbol{Y}|\boldsymbol{X}}$ is not known, i.e., Eq. (2) cannot be evaluated directly. Following [26], it can be shown that a lower bound on $I(\boldsymbol{X}; \boldsymbol{Y})$ is achieved by using mismatched decoding. The samples at the channel output, after fiber transmission, are evaluated as if they were transmitted over an auxiliary channel with transition distribution $q_{\boldsymbol{Y}|\boldsymbol{X}}$ instead of the true yet unknown channel $p_{\boldsymbol{Y}|\boldsymbol{X}}$. Note that $q_{\boldsymbol{Y}|\boldsymbol{X}}$ has the same input and output alphabet as $p_{\boldsymbol{Y}|\boldsymbol{X}}$, i.e., $q_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{x} \in \mathcal{X}} q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \cdot P_{\boldsymbol{X}}(\boldsymbol{x})$. Since MI is an achievable rate, a lower bound on the MI will also be an achievable rate. We denote this lower bound as R and define it as

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}) &= \mathrm{E}_{\boldsymbol{X}\boldsymbol{Y}} \left[\log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{p_{\boldsymbol{Y}}(\boldsymbol{Y})} \right] \\ &\geq \mathrm{E}_{\boldsymbol{X}\boldsymbol{Y}} \left[\log_2 \frac{q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{q_{\boldsymbol{Y}}(\boldsymbol{Y})} \right] \triangleq R. \end{aligned} \qquad (3)$$

Throughout this paper, we define R in units of bit per 4D-symbol (bit/4D-sym). It is apparent that the better $q_{\boldsymbol{Y}|\boldsymbol{X}}$ resembles $p_{\boldsymbol{Y}|\boldsymbol{X}}$, the tighter the bound in (3) will be, and a higher achievable rate is obtained. Although we do not obtain the true MI of the channel, the mismatched decoder approach gives a practical achievable rate since a decoder would also have to assume a channel. Using the auxiliary channel $q_{\boldsymbol{Y}|\boldsymbol{X}}$, a lower bound on MI is estimated in the same fashion as in (2),

$$\frac{1}{N} \sum_{i=1}^{N} \log_2 \frac{q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{x}_i)}{q_{\boldsymbol{Y}}(\boldsymbol{y}_i)} \xrightarrow{p} R, \qquad (4)$$

where $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$, as in (2), are obtained experimentally from the true channel $p_{\boldsymbol{Y}|\boldsymbol{X}}$.

In this paper, we assume $q_{\boldsymbol{Y}|\boldsymbol{X}}$ to be d-dimensional Gaussian distributed,

$$q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{s}_j) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_j)^{\mathrm{T}} \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_j)} \qquad (5)$$

for $j = 1, \ldots, M$ and $\boldsymbol{s}_j$ denoting the jth constellation point. Further, $\boldsymbol{x}$ and $\boldsymbol{y}$ are real d-dimensional column vectors and $|\boldsymbol{\Sigma}_j|$ is the determinant of the covariance matrix $\boldsymbol{\Sigma}_j$. We discuss in detail in Section II-C how the mean values $\boldsymbol{\mu}_j$ and the covariances $\boldsymbol{\Sigma}_j$ of (5) are obtained. The choices for the auxiliary channel in this work are presented in Section II-D.

B. Generalized Mutual Information

In addition to MI, we also compare achievable rates using GMI as a figure of merit. GMI gives an achievable rate for the bit-wise (BW) decoder [29] and has been shown to be an accurate estimate of the post-FEC BER for a wide range of channels, modulation formats, and codes [14], [30]. It should

be noted that, in the same way as for the MI estimates, this is a mismatched decoder approach. While it gives an practically achievable rate, it does not give the highest achievable rate for the BW decoder. However, for decoders applying iterative demapping and decoding [31], [32], it is yet to be validated how accurately GMI represents the post-FEC performance.

Assuming uniformly distributed transmitted symbols, the GMI is estimated as [14]

$$\mathrm{GMI} \approx m - \frac{1}{N} \sum_{k=1}^{m} \sum_{i=1}^{N} \log_2 \left(1 + e^{(-1)^{b_{k,i}} \mathrm{LLR}_{k,i}} \right), \qquad (6)$$

where $b_{k,i}$ is the transmitted bit sequences and $\mathrm{LLR}_{k,i}$ is the log-likelihood ratio with k denoting the bit position and i denoting the ith received symbol. In analogy to the MI computation in (3), the calculation of LLRs also requires an assumption on the underlying channel distribution. In addition to finding the best-possible matching of the auxiliary channel and the true channel distribution, an optimization over a non-negative parameter is in principle required, as discussed in [14, Sec. III-C]. This optimization is omitted in this work, possibly resulting in a lower GMI estimate.

For the general case of a d-dimensional Gaussian distribution as defined in (5), the LLRs are calculated as [14], [33]

$$\begin{aligned} \mathrm{LLR}_{k,i} &= \log \frac{q_{\boldsymbol{Y}|B_k}(\boldsymbol{y}_i|1)}{q_{\boldsymbol{Y}|B_k}(\boldsymbol{y}_i|0)} = \log \frac{\sum\limits_{j:\boldsymbol{s}_j \in \mathcal{X}_1^k} q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{s}_j)}{\sum\limits_{j:\boldsymbol{s}_j \in \mathcal{X}_0^k} q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{s}_j)} \\ &= \log \frac{\sum\limits_{j:\boldsymbol{s}_j \in \mathcal{X}_1^k} \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_j|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{s}_j)^{\mathrm{T}} \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{y}_i - \boldsymbol{s}_j)\right)}{\sum\limits_{j:\boldsymbol{s}_j \in \mathcal{X}_0^k} \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_j|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{s}_j)^{\mathrm{T}} \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{y}_i - \boldsymbol{s}_j)\right)}, \end{aligned} \qquad (7)$$

where $q_{\boldsymbol{Y}|B_k}$ denotes the probability density function of the auxiliary channel conditioned on transmitting the kth bit B_k of $\boldsymbol{X}$. Further, $\mathcal{X}_0^k$ and $\mathcal{X}_1^k$ are the sets of constellations points where the kth bit equals 0 and 1, respectively. It is important to note that, as indicated in Fig. 1, the GMI is dependent on the bit-to-symbol mapping. Throughout this paper, we use Gray-coded constellations in two dimensions.

C. Parameter Estimation

We can see from (5) that the dimensionality and the choice of $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ define the Gaussian auxiliary channel and thus, both MI and GMI. In this work, we differentiate between *static* and *adaptive* mean values. Note for the adaptive case, the mean values are estimated for a batch of received symbols but kept constant over this batch. This should be differentiated from a case where the mean values are tracked over time, which is not in the scope of this paper. The covariances are taken as either independent and identically distributed Gaussian (iidG) or correlated Gaussian (CG).

When static mean values are used, $\boldsymbol{\mu}_j$ is one of the M values of the input alphabet $\mathcal{X}$, i.e., $\boldsymbol{\mu}_j = \boldsymbol{s}_j, \forall j$. For adaptive mean

values, we use the conditional sample mean $\boldsymbol{\mu}_j$ for the jth mean of the multivariate normal distribution,

$$\boldsymbol{\mu}_j = \frac{1}{|\mathcal{I}_j|} \sum_{i \in \mathcal{I}_j} \boldsymbol{y}_i, \tag{8}$$

where the index set $\mathcal{I}_j$ denotes the indices of all $\boldsymbol{y}_i$'s that correspond to a sent constellation point $\boldsymbol{x}_i$ as $\mathcal{I}_j = \{i \in \{1, \ldots, N\} : x_i = s_j\}$.

For iid Gaussian auxiliary channels, $\boldsymbol{\Sigma}_j$ is a $d \times d$ identity matrix I_d multiplied with the average one-dimensional noise variance,

$$\sigma_{\text{1D}}^2 = \frac{1}{d} \sum_{l=1}^{d} \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{y}_{i,l} - \boldsymbol{\mu}_{j,l})^2, \tag{9}$$

where the index l refers to the lth dimension of a d-dimensional vector. Note that σ_{1D}^2 is identical for all j's, i.e., $\boldsymbol{\Sigma}_j = \sigma_{\text{1D}}^2 \cdot I_d, \; \forall j$. In the case of correlated Gaussian auxiliary channels, M $\boldsymbol{\Sigma}_j$'s are calculated, one for each constellation point. The sample covariance conditioned on the jth constellation point being sent is then estimated from the received samples $\boldsymbol{y}$ as

$$\boldsymbol{\Sigma}_j = \frac{1}{|\mathcal{I}_j| - 1} \sum_{i \in \mathcal{I}_j} (\boldsymbol{y}_i - \boldsymbol{\mu}_j)^{\text{T}} (\boldsymbol{y}_i - \boldsymbol{\mu}_j). \tag{10}$$

The estimates of (8) and (10) have the same dimensionality d as the auxiliary channel in (5). It is important to note that this conditional sample covariance means that the noise is not additive since $\boldsymbol{\Sigma}_j$ varies with the transmitted constellation point $\boldsymbol{s}_j$. In other words, the channel noise statistic is conditionally Gaussian.

Throughout this paper, we randomly choose N samples at the output of the DSP to estimate the parameters of $q_{\boldsymbol{Y}|\boldsymbol{X}}$ and N different samples to calculate the achievable rate. For each estimated achievable rate, we use four different experimental batches, from each we take 200000 samples. The final achievable rate is an average over these four batches. This double Monte Carlo approach ensures that we do not overestimate the achievable rate by estimating secondary parameters, i.e., the covariances and the mean values, and our figure of merit, i.e, the achievable rate, from the same sequence.

D. Channel Models

In this work we consider five different models of $q_{\boldsymbol{Y}|\boldsymbol{X}}$ using the Gaussian distributions that are given in Table I. The 2D-iidG distribution is considered as a baseline since this is the most typical assumption made in the literature, i.e., all received constellation points are assumed to have the same variance in both dimensions and no adaptation of the mean values of the received constellation points is applied. We compare this to 2D-CG which assumes different covariances for each constellation point and also the center of each received constellation point is estimated. Since a polarization-multiplexed signal is indeed 4D, and the full 4D-field is required in the DSP, we investigate two 4D distributions. 4D-iidG assumes the same variance for all constellation points and in all four dimensions and 4D-CG estimates a 4D covariance matrix for each constellation point. Both of the 4D estimates applies

TABLE I: Models of the distribution of $q_{\boldsymbol{Y}|\boldsymbol{X}}$

NAME	DESCRIPTION	PARAMETERS	DoFs[§]
1D-iidG	1D iid Gaussian noise in each quadrature and adaptive mean values.[†]	$d = 1$ $\boldsymbol{\mu}_j$ as per (8) $\boldsymbol{\Sigma}_j = \sigma_{\text{1D}}^2 \cdot I_1$	$(4{+}1){\cdot}4 = 20$
2D-iidG	2D iid Gaussian noise in each polarization with static mean values.[‡]	$d = 2$ $\boldsymbol{\mu}_j = \boldsymbol{s}_j$ $\boldsymbol{\Sigma}_j = \sigma_{\text{1D}}^2 \cdot I_2$	$(0{+}1){\cdot}2 = 2$
2D-CG	2D correlated Gaussian (CG) noise and adaptive mean values.[‡]	$d = 2$ $\boldsymbol{\mu}_j$ as per (8) $\boldsymbol{\Sigma}_j$ as per (10)	$(16{\cdot}2{+}16{\cdot}3){\cdot}2 = 160$
4D-iidG	4D iid Gaussian noise and adaptive mean values.	$d = 4$ $\boldsymbol{\mu}_j$ as per (8) $\boldsymbol{\Sigma}_j = \sigma_{\text{1D}}^2 \cdot I_4$	$(16^2 \cdot 4 + 1) = 1025$
4D-CG	4D CG noise and adaptive mean values.	$d = 4$ $\boldsymbol{\mu}_j$ as per (8) $\boldsymbol{\Sigma}_j$ as per (10)	$(16^2 \cdot 4 + 16^2 \cdot 10) = 3584$

[§] The calculated number is the total for all four dimensions of PM-16QAM.
[†] The total R is the sum of the R in all four quadratures.
[‡] The total R is the sum of the R in both polarizations.

adaptation of the center points of the received constellations in 4D. As PM-16QAM can be seen as a 1D modulation format, i.e. 4-ary pulse amplitude modulation in each dimension, we also include a 1D distribution where we assume the same variance for each constellation point and estimate the center of the 1D received constellation points. Shown in Fig. 2 are the received constellations in one polarization with $R \approx 7$ bit/4D-sym for two different experimental configurations, namely 20 Gbaud PM-16QAM in (a)–(b) WDM transmission without inline dispersion compensation and (c)–(d) single-channel transmission with inline dispersion compensation. Indicated in Fig. 2 is the 90 % confidence interval of the variance, calculated from a Chi-square distribution with two free parameters, as black circles for 2D-iidG and white circles for 2D-CG. Green circular markers indicate the transmitted constellation and white square markers the estimated mean values.

Also listed in Table I are the degrees of freedom (DoFs) for each distribution considered for the full 4D signal. The DoFs denote the number of parameters in $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ that are estimated for each auxiliary channel. The expression in the last column of Table I denotes the number of estimated mean values plus the number of covariances, multiplied with $4/d$ to get a 4D expression for the DoFs. For 2D-iidG no mean values are estimated and a single variance is estimated per two dimensions, yielding total of 2 DoFs in four-dimensions. The 1D-iiG estimate with adaptive $\boldsymbol{\mu}_j$'s has 4 mean values and one variance in each of the 4 four dimensions, resulting in 5 DoFs per 1D, or 20 DoFs in 4D. The remaining DoFs are calculated in analogy. Note that for the CG estimates, each $\boldsymbol{\Sigma}_j$ has $d(d+1)/2$ DoFs because $\boldsymbol{\Sigma}_j$ is a symmetric matrix. The number of DoFs is related to the complexity of the receiver. However, the actual complexity will depend on the specific type of decoder that is implemented, which is not in the scope of this paper. To reduce the DoFs, we note that it might be

(a) 2D-iidG : WDM wo. ILDC   (b) 2D-CG : WDM wo. ILDC   (c) 2D-iidG : SC w. ILDC   (d) 2D-CG : SC w. ILDC
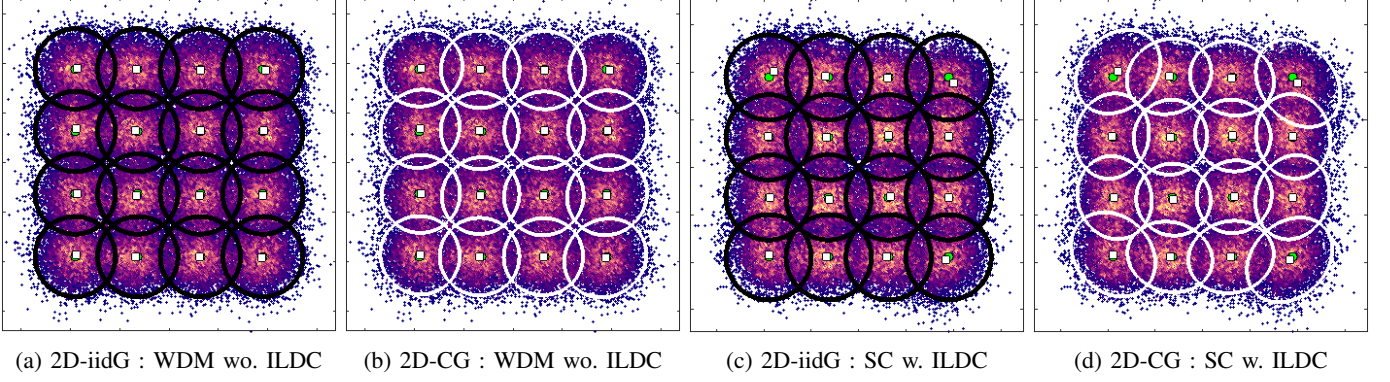
Fig. 2: Measured constellations at $R \sim 7$ bit/4D-sym illustrating the difference between 2D-CG and 2D-iidG estimates showing (a) 2D-iidG and (b) 2D-CG for WDM transmission of 20 Gbaud PM-16QAM without inline dispersion compensation (ILDC). Further shown are (c) 2D-iidG (d) 2D-CG for single channel (SC) transmission of 20 Gbaud PM-16QAM with inline dispersion compensation. Circles and ellipses indicate the 90% confidence interval of the variance for iidG (black) and the covariance for CG (white) estimates, respectively. The green circular markers shows the transmitted constellation which is also what is used for fixed mean values. White square markers show mean values estimates as per (8).

possible to exploit symmetries in Σ_j and $\boldsymbol{\mu}_j$. However, this is beyond the scope of this work.

We also investigate GMI considering both 2D and 4D symbols, denoted as GMI_{2D} and GMI_{4D}, which correspond to the cases of 2D-iidG and 4D-iidG in Table I, although both estimates use fixed mean values which is what is assumed in most conventional bit-wise decoder structures [30]. Thus, GMI_{2D} has 2 DoFs and GMI_{4D} has 1 DoF. However, in Section. V, we discuss the impact of using adaptive mean values for the GMI estimates.

Note that the MI and GMI analysis applies to a discrete-time channel which in this case starts with the digital-to-analog converters (DACs), as shown in Fig. 1. Choices that are made in the transmitter (Tx) on pulse-shaping or dispersion pre-compensation will have an influence on the achievable rate since it actually changes the discrete-time channel. The same goes for the receiver (Rx) side, where the choices of algorithms will influence the achievable rate. An example that affects the results significantly is if any nonlinear compensation technique is used or not. More obvious perhaps is the impact from choices made on the actual transmission channel such as amplifier spacing, optical launch power, amplifier technology, and whether inline dispersion compensation is used or not.

III. EXPERIMENTAL SETUP

The transmitter is shown in Fig. 3a. The electrical driving signals are generated using a 4-channel arbitrary waveform generator (AWG) running at 60 Gs/s, producing either 10 Gbaud or 20 Gbaud PM-16QAM signals with root-raised cosine (RRC) pulses using a roll-off factor of 0.5. The AWG drives two IQ modulators that are modulating in total 7 WDM channels using distributed feedback lasers (DFBs) with $\sim$150 kHz linewidth as sources. For single-channel transmission, all lasers except the center channel are turned off. The IQ modulators are modulating the even and odd WDM channels separately and the signals are decorrelated in the AWG. After the IQ modulators, the optical signals are combined and sent to a polarization-multiplexing emulation stage based on split, delay, and recombination with orthogonal polarization states.

The optical signals are propagated over a recirculating loop shown in Fig. 3b. The loop consists of two spans of 80 km of conventional single-mode fiber (SMF) amplified by erbium-doped fiber amplifiers (EDFAs). Before the first span, a variable band-pass filter (V-BPF) is used to filter out excessive amplified spontaneous emission (ASE) noise. The bandwidth of the filter is varied depending on the type of signal transmitted. Before the second span, a programmable wavelength-selective switch is used both as V-BPF and, when
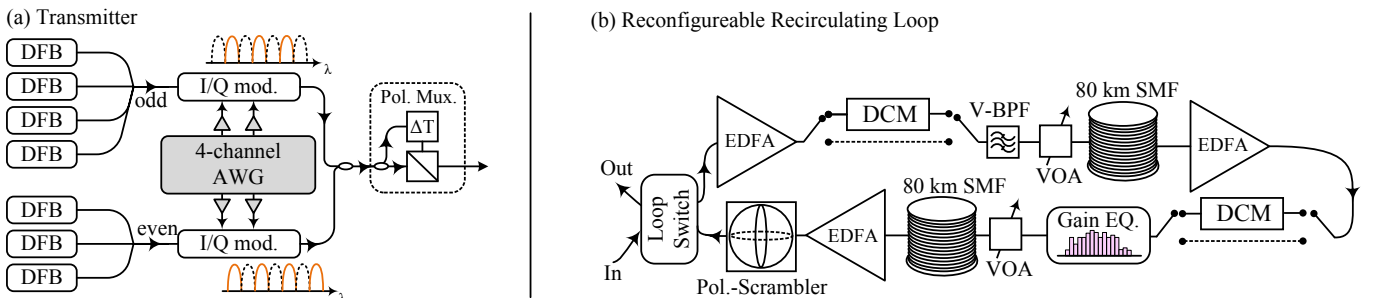

Fig. 3: (a) Schematics of the transmitter (b) Recirculating loop with 80 km spans. Abbreviations are explained in the text.
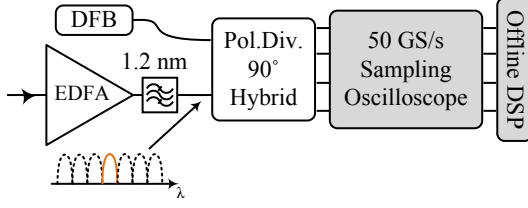
Fig. 4: Schematics of the coherent receiver.

WDM-signals are transmitted, as a gain equalizer. Preceding both spans are variable optical attenuators (VOAs) that control the optical launch power. Also in the loop is a polarization scrambler (pol.-scrambler) that is synchronized to the round-trip time of the loop and a third EDFA compensates for the loss of the pol.-scrambler and the loop-switching components. The recirculating loop can be configured to have either inline dispersion compensation using dispersion-compensating modules (DCMs) based on fiber-Bragg gratings [34] or uncompensated by simply bypassing the DCMs. The DCMs are placed directly after the EDFAs that are preceding the fiber spans, as no extra EDFAs are then needed. It is important to note that the DCMs themselves inflict no nonlinear distortions [35].

The center channel is detected using a coherent receiver based on an integrated polarization-diverse optical hybrid with balanced photo-detectors as depicted in Fig. 4. As local-oscillator (LO) a DFB laser of the same type as the transmitter lasers is used. The electrical signals are sampled using a 50 GS/s real-time oscilloscope with 33 GHz bandwidth and processed using off-line DSP.

A. Digital Signal Processing

The DSP starts with optical front-end compensation followed by resampling to 2 samples/symbol. If there is no inline dispersion compensation, electronic dispersion compensation (EDC) implemented in the frequency domain is applied. Polarization demultiplexing and adaptive equalization are applied using four FIR filters in a butterfly structure. The filter taps are initially updated using the constant-modulus algorithm (CMA) for pre-convergence followed by decision-directed least mean square (DD-MLS) for final adaptation. The number of taps is either 17 for 20 Gbaud signals or 25 for 10 Gbaud. Frequency estimation based on the fast Fourier transform and carrier phase estimation (CPE) based on blind phase search with 32 test angles [36] are performed within the DD-LMS loop. The block length of the phase tracking is either 128 for 20 Gbaud signals or 256 for 10 Gbaud signals.

As mentioned previously, it should be noted that for the achievable rate estimates, the DSP is part of the channel and can affect the estimates. The adaptive equalizer and the phase tracking algorithms assume additive white Gaussian noise statistics of the noise which, as is explained in the next section, is not the case for all of the systems investigated. Instead of using a blind receiver, a pilot-symbol based approach could be used [37]. This type of receiver poses an interesting question of the trade-off between pilot overhead and FEC overhead. The most notable part of the DSP that influences the channel is
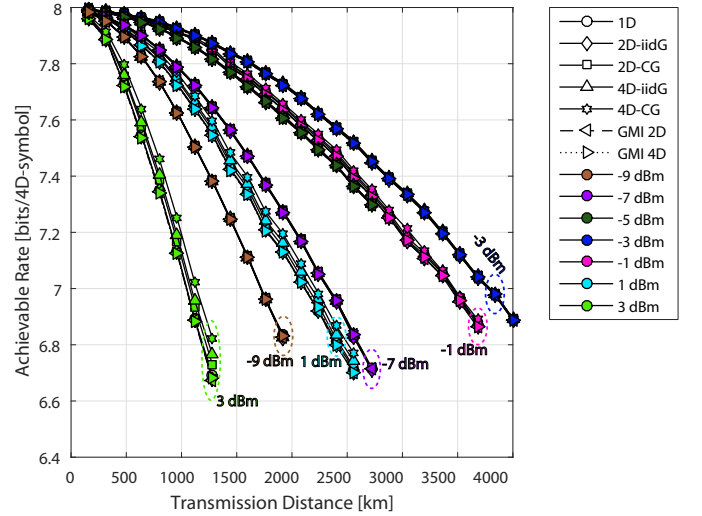

Fig. 5: WDM transmission of 20 Gbaud PM-16QAM with 30 GHz channel separation, without inline dispersion compensation.

the phase tracking. For instance, we note that the block length of the CPE influences the nonlinear memory of the channel [38]. Further, we limit ourself to not incorporating digital back propagation (DBP) in the DSP used in this study. It is known that DBP can greatly reduce the nonlinear distortions [39] at the cost of increased complexity. It has been shown that the achievable rate when DBP is used is dependent on the number of channels that are considered [40], and that its performance is dependent on frequency stabilization between the WDM channels [41]. Although interesting, we leave DBP for a future study.

IV. EXPERIMENTAL RESULTS

In the following section, the achievable rates for different link configurations are presented. We investigate single channel transmission and WDM transmission of 10 Gbaud and 20 Gbaud PM-16QAM with and without inline dispersion compensation.

A. WDM transmission of 20 Gbaud PM-16QAM, no inline dispersion compensation

In recent times, the majority of the fiber optical communication experiments reported are over transmission links without inline dispersion compensation. The main reason for this is that by accumulating the dispersion, in general the nonlinear impairments have less impact on the achievable transmission distances. Shown in Fig. 5 is the achievable rate using the different estimates as a function of transmission distance for different launch powers for WDM transmission of 20 Gbaud PM-16QAM with 30 GHz channel spacing. The main finding here is that the different assumptions on the distributions of the achievable rate estimates have a negligible difference at the optimal launch power of -3 dBm. This is also true for lower launch power and slightly higher launch power. It is only in the extreme case of 1 dBm and 3 dBm launch power that
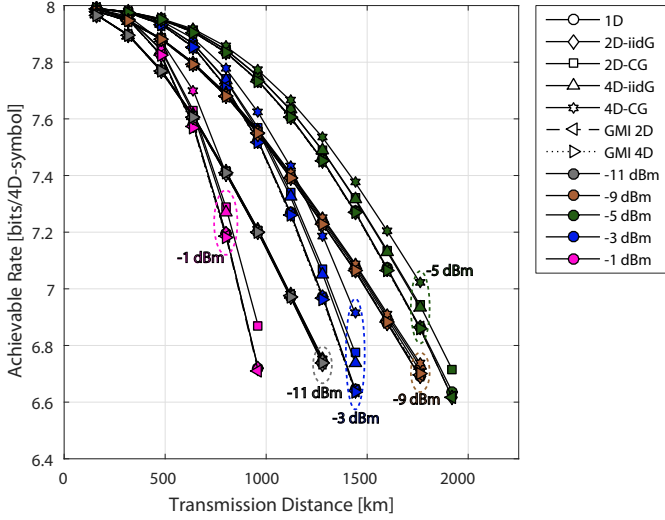
Fig. 6: WDM transmission of 20 Gbaud PM-16QAM with 30 GHz channel separation, with inline dispersion compensation.

Fig. 7: Single channel transmission of 20 Gbaud PM-16QAM without inline dispersion compensation.

significant difference between the estimates can be observed. At 1280 km and 3 dBm launch power 2D-GMI, 4D-GMI, 1D-iddG, and 2D-iidG give roughly the same achievable rate. 2D-CG and 4D-iidG see a small gain over 2D-iidG and the highest gain is seen by 4D-CG which has 0.13 bit/4D-symb higher achievable rate than the lowest cases. However, note that this is an unrealistically high launch power as the transmission distance is only 30 % of that achieved with the optimal power. We may thus conclude that for WDM transmission links without inline dispersion compensation, the memoryless noise statistics is iid Gaussian and a decoder working in 2D-iidG or even 1D-iidG should achieve as high rate as a receiver working with the more complex noise statistics. This is in good agreement with [25] where we showed that the memoryless 2D-iidG channel distribution assumption is valid for 28 Gbaud PM-QPSK and polarization-switched QPSK in (WDM) transmission without inline dispersion compensation.

B. WDM transmission of 20 Gbaud PM-16QAM, with inline dispersion compensation

Fig. 6 shows WDM transmission of 20 Gbaud PM-16QAM signals with a channel spacing of 30 GHz. For this case, inline dispersion compensation is used. Note that for clarity purpose, the results for -7 dBm have been removed as this curve is similar to that of -5 dBm. We note that for the optimal launch power of -5dBm, the achievable transmission distance is severely reduced for this scenario compared to the previous case without inline dispersion compensation. Compared to the non-dispersion managed link with the same WDM signals, at an achievable rate of 7 bit/4D-sym, the transmission distance is roughly halved. We note that for very low launch powers, where nonlinear distortions can be neglected, there is no apparent difference between any of the estimates. However, for the optimal launch power there is a gain in the achievable rate by considering more sophisticated distributions. For this
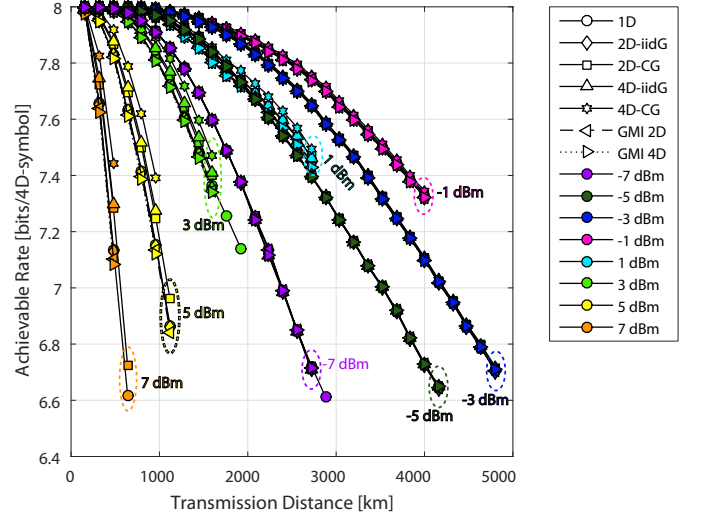
launch power, the difference between 2D-GMI, 4D-GMI, 1D-iidG, and 2D-iidG is insignificant. However, 4D-iidG and 2D-CG see a small gain, for instance at a distance of 1760 km the achievable rate is roughly 0.08 bit/4D-sym higher than the previously mentioned estimates. The largest gain is seen by 4-CG which has a 0.16 bit/4D-sym higher achievable rate compared to the lowest estimates. As the launch power increases, the gain seen by the 4D-CG estimate increases.

C. Single-channel transmission of 20 Gbaud PM-16QAM, without inline dispersion compensation

Shown in Fig. 7 are the achievable rates for all estimates as a function of transmission distance for different launch power for single-channel 20 Gbaud PM-16QAM without inline dispersion compensation. We notice that compared to the WDM case for the same link, the optimal launch power is increased by 2 dB. For low launch powers, there is no distinctive difference between the different estimates. In the same way as for the WDM case without inline dispersion compensation, for the highest launch powers, a difference can be observed. However, in this case the difference is significant already for launch powers that are close to the optimal and even for the optimal launch power a small gain can be observed for the 4D-CG estimate.

This means that for a single-channel system, even without inline dispersion compensation, the noise statistics are no longer iid Gaussian, especially for launch powers higher than the optimal. However, since at the optimal launch power the difference in achievable rate for the 4D-CG estimate and the 2D-iidG is less than 0.02 bit/4D-sym at 4000 km, it can be argued that even for this scenario, assuming iid Gaussian statistics of the channel transition distribution is a reasonable trade-off between achievable rate and complexity.
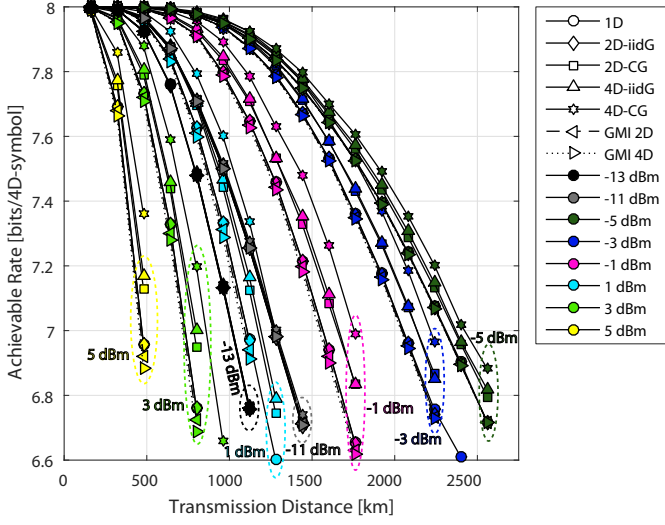
Fig. 8: Single channel transmission of 20 Gbaud PM-16QAM with inline dispersion compensation.

Fig. 9: WDM transmission of 10 Gbaud PM-16QAM with 15 GHz channel separation, without dispersion compensation.

D. Single-channel transmission of 20 Gbaud PM-16QAM, with inline dispersion compensation

Presented in Fig. 8 is the achievable rate for single-channel 20 Gbaud PM-16QAM with inline dispersion compensation. Note that the results for -7 dBm launch power has been omitted for clarity. The optimal launch power is -5 dBm, which is the same as for the WDM transmission over the same link. Compared to the single-channel transmission of 20 Gbaud PM-16QAM without inline dispersion compensation, we note that the optimal launch power is 4 dB lower. We also note that for this link, there is a big difference between the different estimates at the optimal launch power. At 2400 km, the 4D-CG estimate has a gain in achievable rate of roughly 0.13 bit/4D-sym over 2D-iidG. Further, 4D-iidG and 2D-CG see an intermediate gain of 0.07 bit/4D-sym and 0.06 bit/4D-sym, respectively. For this link, the noise statistics are not iid Gaussian except for low launch powers. For the optimal launch power, compared at an achievable rate of 7 bit/4D-symb, 5% increased transmission distance can be achieved by assuming 4D-CG instead of 2D-iidG statistics.

E. WDM transmission of 10 Gbaud PM-16QAM, without inline dispersion compensation

In this part, the symbol rate is decreased to 10 Gbaud, while the SE is kept constant by changing the channel spacing to 15 GHz. Shown in Fig. 9 is the achievable rate as a function of transmission distance in this case and for the link without inline dispersion compensation. For the optimal launch power of -7 dBm we note, opposed to the 20 Gbaud scenario over the same link, that a difference can be observed between the different estimates. At 4000 km, 2D-CG and 4D-iidG see a gain of roughly 0.03 bit/4D-sym over 2D-iidG. 4D-CG sees the largest gain of 0.04 bit/4D-sym. While these gains are not large, we however note that even if the WDM link is without inline dispersion compensation, if the symbol rate is low enough, the memoryless statistic of the received signal is not perfectly described by 2D-iidG distributions.
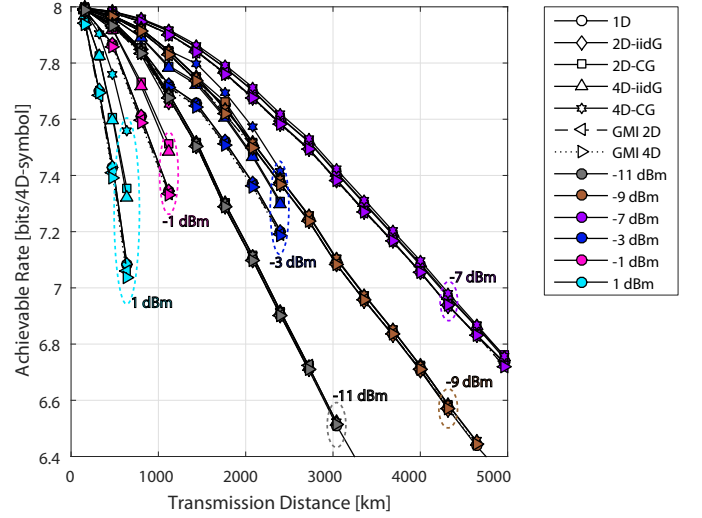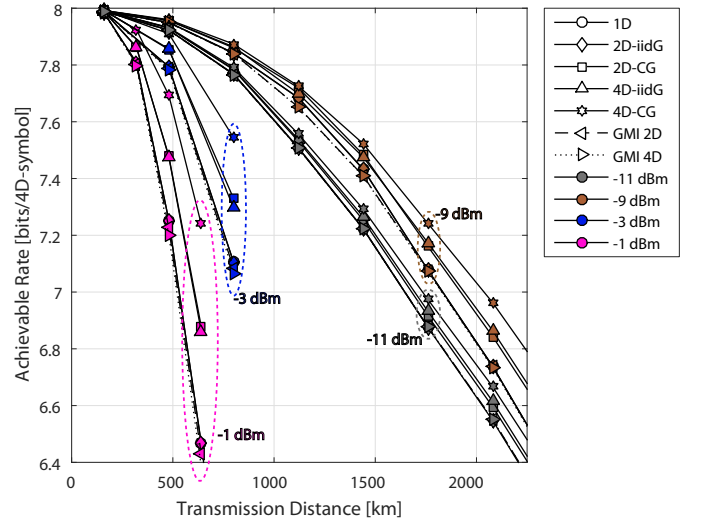

Fig. 10: WDM transmission of 10 Gbaud PM-16QAM with 15 GHz channel separation, with inline dispersion compensation.

F. WDM transmission of 10 Gbaud PM-16QAM, with inline dispersion compensation

Depicted in Fig. 10 is the achievable rate for WDM transmission with 15 GHz channel spacing for 10 Gbaud PM-16QAM with inline dispersion compensation. Some measured launch powers are not plotted for clarity reasons. Note that this case has the same SE as the 20 Gbaud case with 30 GHz channel spacing (Fig. 6). For the optimal launch power there is a clear difference between the different estimates. At 1760 km, 2D-CG sees a gain of 0.07 bit/4D-sym and 4D-iidG have a gain of 0.10 bit/4D-sym. 4D-CG has the largest gain of 0.17 bit/4D-sym.

V. DISCUSSION

First and foremost, we note that for the most realistic scenario, i.e. 10 or 20 Gbaud WDM transmission without any

TABLE II: Transmission distance difference in percentage compared to the commonly used 2D-iidG distribution at 7 bit/4D-symb.

System Setup	1D-iidG	2D-CG	4D-iidG	4D-CG	2D-GMI	4D-GMI
WDM 20 Gbaud PM-16QAM, w/o ILDC	$\sim 0\%$	$\sim 0\%$	$\sim 0\%$	$\sim 0\%$	$\sim 0\%$	$\sim 0\%$
WDM 20 Gbaud PM-16QAM, w ILDC	$\sim 0\%$	1.8%	2.2%	3.9%	$\sim 0\%$	$\sim 0\%$
SC 20 Gbaud PM-16QAM, w/o ILDC†	$\sim 0\%$	0.5%	0.9%	1.6%	$\sim 0\%$	$\sim 0\%$
SC 20 Gbaud PM-16QAM, w ILDC	0.2%	2.5%	2.8%	5.1%	$\sim 0\%$	$\sim 0\%$
WDM 10 Gbaud PM-16QAM, w ILDC	$\sim 0\%$	4.6%	5.9%	11.2%	$\sim 0\%$	$\sim 0\%$
WDM 10 Gbaud PM-16QAM, w/o ILDC	$\sim 0\%$	1.9%	2.3%	2.9%	$\sim 0\%$	$\sim 0\%$

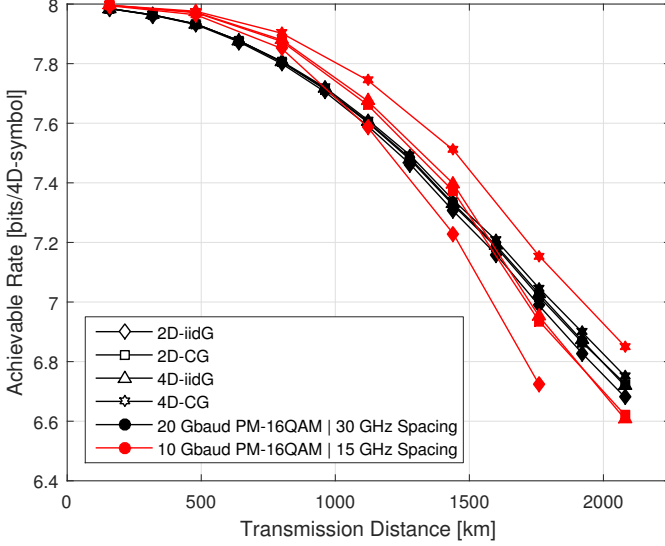† Measured at 7.35 bit/4D-symb instead of 7.00 bit/4D-symb.

Fig. 11: Comparison between the achievable rates for two WDM scenarios with the same SE, namely, 20 Gbaud PM-16QAM on 30 GHz spacing and 10 Gbaud PM-16QAM on 15 GHz spacing for the link with inline dispersion compensation.

Fig. 12: Comparison between 2D-iidG estimate with and without adaptive mean values of the constellation points.

inline dispersion compensation, it is a good approximation to assume 2D iid Gaussian noise distributions, which is most commonly done in decoders. This is true also for single-channel transmission over links without inline dispersion compensation. Presented in Table II is the approximate gain in transmission distance over a 2D-iidG channel distribution assumption, for the optimal launch power at 7 bit/4D-sym for all the different transmission scenarios considered in this paper. We note that for the scenarios considered here, there is never any significant difference between the 2D-iidG, 2D-GMI, 4D-GMI, and 1D-iidG channel assumptions.

An interesting comparison is the WDM transmission of 20 Gbaud PM-16QAM with 30 GHz channel separation and 10 Gbaud PM-16QAM with 15 GHz spacing, as these two cases have the same spectral efficiency. Shown in Fig. 11 are these two cases for the 2D-iidG, 2D-CG, 4D-iidG, and 4D-CG estimates at the optimal launch power. We first note that for the 2D-iidG estimate, the longest transmission distance is achieved by the 20 Gbaud case for achievable rates below 7.7 bit/4D-symb. We also note that the difference between the estimates is small in the 20 Gbaud case while for the 10 Gbaud
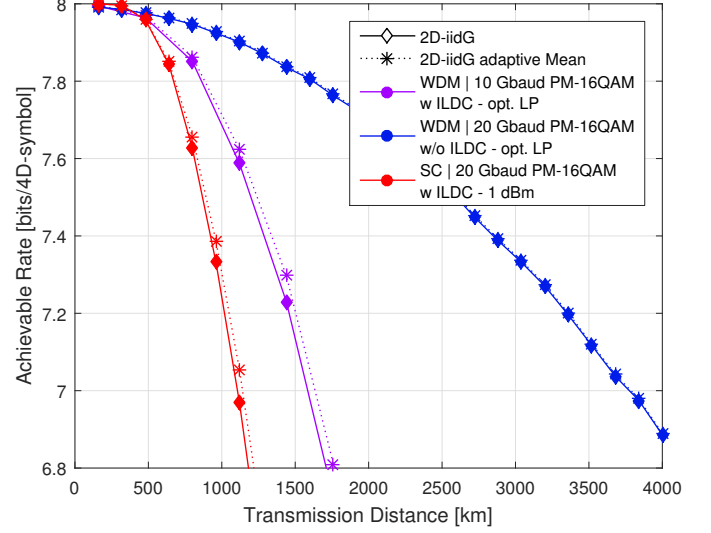
case, a large difference is observed. The 2D-CG and 4D-iidG estimates for 10 Gbaud see roughly the same achievable rates as for the 20 Gbaud case. For high achievable rates, the 10 Gbaud system achieves a longer transmission distance compared to the 20 Gbaud case for any estimate. This is most likely due to the fact that the number of channels are kept the same, i.e. the spectral width that the 10 Gbaud case occupies is half of that of the 20 Gbaud case. More notably though is that the 4D-CG estimate for 10 Gbaud achieves the longest transmission distance for any achievable rate. We conclude that for the system with inline dispersion compensation, there seems to be different optimal symbol rates for a fixed SE, depending on which distribution is assumed in the decoder. This is a similar finding to that of [42], [43] where it is shown that the nonlinear interference can be reduced by dividing a signal into lower symbol rate subcarriers. However, a more detailed study on several different symbol rates is required to fully understand this effect. One possible explanation for the higher achievable rate for the 10 Gbaud case is that all the estimates in this paper are memoryless and in the 20 Gbaud case the dispersion increases the nonlinear memory of the channel more rapidly than in the 10 Gbaud case, hence there is more loss of information for the 20 Gbaud signal. We also investigated this for a system without inline compensation and
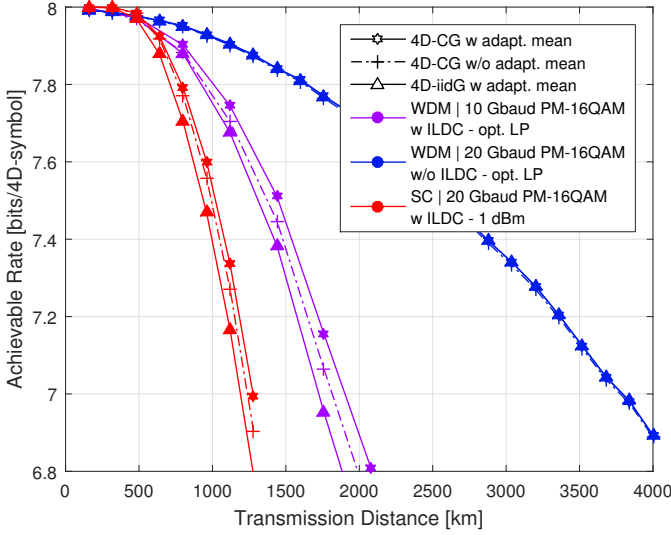
Fig. 13: Comparison between 4D-CG with and without adaptive mean values for the 4D-constellation points and 4-iidG with adaptive means.

Fig. 14: Achievable rates for 4D and 2D GMI, with and without adaptive mean values of the received constellation.

no such effect is then observed.

The baseline measure in this paper is 2D-iidG with non-adaptive means of the constellation points. We note that for a nonlinear channel, some of the gain for the more complex estimates are indeed from this adaptation and not only due to the more sophisticated variance estimates. In Fig. 12, 2D-iidG with fixed means is compared to the same case but with adaptive means for three different transmission scenarios. We note that for the optimal launch power for WDM transmission of 20 Gbaud PM-16QAM without inline dispersion compensation there is no distinctive difference, which is expected as we saw no significant difference for any estimate for this case (Fig. 5). For the optimal launch power of WDM transmission of 10 Gbaud PM-16QAM in the link with inline dispersion compensation, a small gain in the achievable rate can be observed. The same is true for the high launch power of 1 dBm for single-channel transmission of 20 Gbaud PM-16QAM over the same link.

In Fig. 13 we compare the best performing estimate, 4D-CG, with and without adaptive mean values of the received constellation points for the same cases as for 2D-iidG in Fig. 12. We also added 4D-iidG with adaptive means in this comparison. Again, there is no difference for the WDM transmission without inline dispersion compensation. However, for the two cases with inline dispersion compensation we can indeed conclude that some of the gain for this format comes from the adaptation of the constellation points. However, if we compare to the 4D-iidG with adaptive means, where each 4D-constellation point is considered to have the same variance in all dimensions, we can see that the achievable rate is considerable lower than the 4D-CG case without adaptive means. Hence, we can draw the conclusions that the gain seen for the 4D estimates is dependent on both the fact that the means are adapted in a 4D space and on the 4D estimates of the variances.

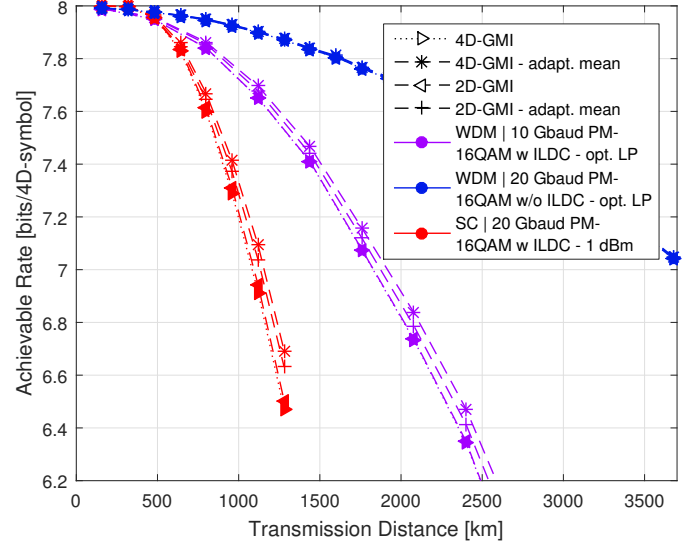Another interesting comparison regarding the adaptive mean

of the received constellation points concerns the GMI estimates. In this paper we have compared different MI estimates to 2D and 4D GMI estimates, where we use a fixed received constellation. The reason for this is that it corresponds to the most conventional decoder structures. However, in Fig. 14 we compare the 2D and 4D GMI estimates to the case where we allow adaptive estimation of the received constellation points for WDM transmission of 20 Gbaud PM-16QAM without inline dispersion compensation at the optimal launch power, WDM transmission of 10 Gbaud PM-16QAM with inline dispersion compensation at the optimal launch power, and single-channel transmission of 20 Gbaud PM-16QAM at a higher-than-optimal launch power of −1 dBm. For WDM transmission of 20 Gbaud PM-16QAM without inline dispersion compensation, no difference is seen between the compared cases. For the systems without inline dispersion compensation with either 10 Gbaud WDM transmission or 20 Gbaud single-channel transmission, we note that there is no significant difference between 2D and 4D GMI without adaptive mean values. However, the 4D-GMI estimate with adaptive mean values sees a clear gain over the other GMI estimates for these two cases. We also note that the 2D-GMI with adaptive means sees an intermediate gain between not estimating the mean values and estimating the mean values in four dimensions. At 2400 km, 4D-GMI with adaptive means sees a gain of 0.13 bit/4D-sym over 2D-GMI without adaptive means. This shows that for the bit-wise decoder, for certain transmission systems such as the WDM transmission of 10 Gbaud PM-16QAM over transmission link with inline dispersion compensation, there are possible gains by designing decoders that can estimate the mean values of the received constellation points in four dimensions. We note that this should be a minor tweak to existing 4D decoders [30], [44].

VI. Conclusions

We have experimentally investigated the achievable rate using GMI and MI with different assumptions on the channel distribution for single-channel and WDM transmission of 10 Gbaud and 20 Gbaud PM-16QAM signals for transmission links with and without inline dispersion compensation. We have shown that for most practical scenarios, assuming that the transmission channel has independent and identically distributed Gaussian distribution of the noise in each dimension is a good approximation. In other words, for a practical system, decoders that are assuming fixed constellation means and the same variance for all constellation points have no significant penalty over decoders using more sophisticated distributions. However, for systems with inline dispersion compensation, we show that there is gain in using 4D distributions, most notably using 4D correlated Gaussian distributions with adaptive mean values which shows a small but significant gain even at the optimal launch power. We also show that for all cases, the difference between GMI and MI estimates using 2D-iidG or 1D-iidG distributions is negligible. For the more extreme cases, for instance single-channel transmission with high launch powers, the assumption that the channel is Gaussian with the same variance in each dimension is no longer valid and large gains are seen by assuming 4D correlated Gaussian distributions.

Acknowledgments

References

[1] S. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1164–1179, 2010.

[2] R. Essiambre and R. Tkach, "Capacity trends and limits of optical communication networks," *Proceedings of the IEEE*, vol. 100, no. 5, pp. 1035–1055, 2012.

[3] H. Zhang, J.-X. Cai, H. G. Batshon, C. R. Davidson, Y. Sun, M. Mazurczyk, D. G. Foursa, A. Pilipetskii, G. Mohs, and N. S. Bergano, "16QAM transmission with 5.2 bits/s/Hz spectral efficiency over transoceanic distance," *Optics Express*, vol. 20, no. 11, pp. 11 688–11 693, 2012.

[4] J.-X. Cai, H. G. Batshon, H. Zhang, M. Mazurczyk, O. Sinkin, D. G. Foursa, A. Pilipetskii, G. Mohs, and N. S. Bergano, "30.4 Tb/s transmission over transpacific distance using 200 Gb/s and dual wavelength 400 Gb/s 16QAM at 6.0 b/s/Hz spectral efficiency," *Optics Express*, vol. 22, no. 8, pp. 9116–9122, 2014.

[5] D. Qian, M.-F. Huang, S. Zhang, Y. Zhang, Y.-K. Huang, F. Yaman, I. B. Djordjevic, and E. Mateo, "30Tb/s C- and L-bands bidirectional transmission over 10,181km with 121km span length," *Optics Express*, vol. 21, no. 12, pp. 14 244–14 250, 2013.

[6] K. Igarashi, T. Tsuritani, I. Morita, Y. Tsuchida, K. Maeda, M. Tadakuma, T. Saito, K. Watanabe, K. Imamura, R. Sugizaki, and M. Suzuki, "Super-Nyquist-WDM transmission over 7,326-km seven-core fiber with capacity-distance product of 1.03 Exabit/s·km," *Optics Express*, vol. 22, no. 2, pp. 1220–1228, 2014.

[7] ITU-T, "Interfaces for the optical transport network," *ITU-T G.975*, 2000.

[8] ITU-T, "Forward error correction for high bit-rate DWDM submarine systems," *ITU-T G.975.1*, 2004.

[9] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[10] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[11] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1261–1271, 1996.

[12] T. Mizuochi, "Next Generation FEC for Optical Communication," in *Proc. Optical Fiber Communication Conference (OFC)*, 2008, paper OTuE5.

[13] A. Leven, F. Vacondio, L. Schmalen, S. Brink, and W. Idler, "Estimation of soft FEC performance in optical transmission experiments," *Photonics Technology Letters*, vol. 23, no. 20, pp. 1547–1549, 2011.

[14] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems," *Journal of Lightwave Technology*, vol. 33, no. 20, pp. 4338–4352, 2015.

[15] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *Journal of Lightwave Technology*, vol. 23, no. 11, pp. 3755–3763, 2005.

[16] R. Essiambre, G. Kramer, P. Winzer, G. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010.

[17] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *Journal of Lightwave Technology*, vol. 31, no. 23, pp. 3839–3852, 2013.

[18] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik, "On achievable rates for long-haul fiber-optic communications," *Optics Express*, vol. 23, no. 7, pp. 9183–9191, 2015.

[19] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," *Journal of Lightwave Technology*, vol. 32, no. 16, pp. 2862–2876, 2014.

[20] G. Kramer, M. Yousefi, and F. Kschischang, "Upper bound on the capacity of a cascade of nonlinear and noisy channels," in *Proc. Information Theory Workshop (ITW)*, 2015.

[21] D. Millar, R. Maher, D. Lavery, T. Koike-Akino, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel, "Transceiver-limited high spectral efficiency Nyquist-WDM systems," in *Proc. Optical Fiber Communication Conference (OFC)*, 2015, paper Th2A.13.

[22] R. Maher, D. Lavery, D. Millar, A. Alvarado, K. Parsons, R. Killey, and P. Bayvel, "Reach enhancement of 100% for a DP-64QAM super-channel using MC-DBP," in *Proc. Optical Fiber Communication Conference (OFC)*, 2015, paper Th4D.5.

[23] R. Maher, A. Alvarado, D. Lavery, and P. Bayvel, "Modulation order and code rate optimisation for digital coherent transceivers using generalised mutual information," in *Proc. European Conference on Optical Communication (ECOC)*, 2015, paper Mo.3.3.4.

[24] D. Marsella, M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *Journal of Lightwave Technology*, vol. 32, no. 5, pp. 908–916, 2014.

[25] T. A. Eriksson, T. Fehenberger, N. Hanik, P. A. Andrekson, M. Karlsson, and E. Agrell, "Four-dimensional estimates of mutual information in coherent optical communication experiments," in *Proc. European Conference on Optical Communication (ECOC)*, 2015, paper We.4.6.5.

[26] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, 2006.

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

[28] S. Verdú and T. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[29] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[30] A. Alvarado and E. Agrell, "Four-dimensional coded modulation with bit-wise decoders for future optical communications," *Journal of Lightwave Technology*, vol. 33, no. 10, pp. 1993–2003, 2015.

[31] I. Djordjevic, M. Cvijetic, L. Xu, and T. Wang, "Using LDPC-coded modulation and coherent detection for ultra highspeed optical transmission," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3619–3625, 2007.

[32] H. Bülow and E. Masalkina, "Coded modulation in optical communications," in *Proc. Optical Fiber Communication Conference (OFC)*, 2011, paper OThO1.

[33] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2756–2765, 2009.

[34] Proximion AB, "Dispersion Management - White paper fiber Bragg grating based DCM," http://www.proximion.com/dispersion-compensation, 101006-F, 2013.

[35] E. Tipsuwannakul, J. Li, T. A. Eriksson, L. Egnell, F. Sjöström, J. Pejnefors, P. A. Andrekson, and M. Karlsson, "Influence of fiber-Bragg grating-onduced group-delay ripple in high-speed transmission systems," *Journal of Optical Communications and Networking*, vol. 4, no. 6, pp. 514–521, 2012.

[36] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations," *Journal of Lightwave Technology*, vol. 27, no. 8, pp. 989–999, 2009.

[37] R. Elschner, F. Frey, C. Meuer, J. K. Fischer, S. Alreesh, C. Schmidt-Langhorst, L. Molle, T. Tanimura, and C. Schubert, "Experimental demonstration of a format-flexible single-carrier coherent receiver using data-aided digital signal processing," *Optics Express*, vol. 20, no. 27, pp. 28 786–28 791, 2012.

[38] T. Fehenberger, N. Hanik, T. A. Eriksson, P. Johannisson, and M. Karlsson, "On the impact of carrier phase estimation onphase correlations in coherent fiber transmission," in *Proc. Tyrrhenian International Workshop on Digital Communications*, 2015, paper P3.2.

[39] E. Ip and J. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, 2008.

[40] T. Fehenberger and N. Hanik, "Digital back-propagation of a superchannel: Achievable rates and adaption of the GN model," in *Proc. European Conference on Optical Communication (ECOC)*, 2014, paper We.3.3.6.

[41] E. Temprana, E. Myslivets, B.-P. Kuo, L. Liu, V. Ataie, N. Alic, and S. Radic, "Overcoming Kerr-induced capacity limit in optical fiber transmission," *Science*, vol. 348, no. 6242, pp. 1445–1448, 2015.

[42] P. Poggiolini, Y. Jiang, A. Carena, G. Bosco, and F. Forghieri, "Analytical results on system maximum reach increase through symbol rate optimization," in *Proc. Optical Fiber Communication Conference (OFC)*, 2015, paper Th3D.6.

[43] A. Carbo Meseguer, J. Renaudier, R. Rios-Müller, P. Tran, and G. Charlet, "Impact of bandwidth efficiency in nonlinear tolerance of multicarrier modulations," *Journal of Lightwave Technology*, vol. PP, no. 99, 2015.

[44] H. G. Batshon, I. Djordjevic, and T. Schmidt, "Ultra high speed optical transmission using subcarrier-multiplexed four-dimensional LDPC-coded modulation," *Optics Express*, vol. 18, no. 19, pp. 20 546–20 551, 2010.