CHALMERS
UNIVERSITY OF TECHNOLOGY

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, Oct. 2015; reprinted, vol. 34, no. 2, pp. 707–721, Jan. 2016.

Replacing the Soft-decision FEC Limit Paradigm in the Design of Optical Communication Systems

Alex Alvarado, *Senior Member, IEEE*, Erik Agrell, *Senior Member, IEEE*, Domaniç Lavery, *Member, IEEE*, Robert Maher, *Senior Member, IEEE*, and Polina Bayvel, *Fellow, IEEE*

(Invited Paper)

***Abstract*—The FEC limit paradigm is the prevalent practice for designing optical communication systems to attain a certain bit-error rate (BER) without forward error correction (FEC). This practice assumes that there is an FEC code that will reduce the BER after decoding to the desired level. In this paper, we challenge this practice and show that the concept of a channel-independent FEC limit is invalid for soft-decision bit-wise decoding. It is shown that for low code rates and high order modulation formats, the use of the soft-decision FEC limit paradigm can underestimate the spectral efficiencies by up to 20%. A better predictor for the BER after decoding is the generalized mutual information, which is shown to give consistent post-FEC BER predictions across different channel conditions and modulation formats. Extensive optical full-field simulations and experiments are carried out in both the linear and nonlinear transmission regimes to confirm the theoretical analysis.**

I. Introduction and Motivation

Forward error correction (FEC) and multilevel modulation formats are key technologies for realizing high spectral efficiencies in optical communications. The combination of FEC and multilevel modulation is known as coded modulation (CM), where FEC is used to recover the sensitivity loss from the nonbinary modulation. While in the past optical communication systems were based on hard-decision (HD) FEC, modern systems use *soft-decision* FEC (SD-FEC).

Current digital coherent receivers are based on powerful digital signal processing (DSP) algorithms, which are used to detect the transmitted bits and to compensate for channel impairments and transceiver imperfections. The optimal DSP should find the most likely *coded sequence*. However, this is hard to realize in practice, and thus, most receivers are implemented suboptimally. In particular, detection and FEC decoding are typically decoupled at the receiver: soft information on the code bits is calculated first, and then, an SD-FEC decoder is used. We refer to this receiver structure as a *bit-wise* (BW) decoder, also known in the literature as a bit-interleaved coded modulation (BICM) receiver [1], [2], owing its name to the original works [3], [4], where a bit-level interleaver was included between the FEC encoder and mapper. In the context of optical communications, BW decoders have been studied, e.g., in [5]–[10].

An alternative to BW decoders is to use iterative demapping (ID) and decoding, i.e., when the FEC decoder and demapper exchange soft information on the code bits iteratively. This structure is known as BICM-ID and was introduced in [11]–[13]. BICM-ID for optical communications has been studied in [14]–[16], [7, Sec. 3], [17, Sec. 3], [18, Sec. 4]. Due to the inherent simplicity of the (noniterative) BW receiver structure, BICM-ID is not considered in this paper.

For simplicity, researchers working on optical communications typically use offline DSP. In this case, and to meet higher-layer quality of service requirements, the bit-error rate (BER) after FEC decoding—in this paper referred to as post-FEC BER or $\mathrm{BER}_{\mathrm{post}}$—should be as low as 10^{-12} or 10^{-15}. Since such low BER values cannot be reliably estimated by Monte-Carlo simulations, the conventional design strategy has been to simulate the system without FEC encoding and decoding, and optimize it for a much higher BER value, the so-called "FEC limit" or "FEC threshold". The rationale for this approach, which we call the *FEC limit paradigm*, is that a certain BER without coding—here referred to as pre-FEC BER or $\mathrm{BER}_{\mathrm{pre}}$—supposedly can be reduced to the desired post-FEC BER by previously verified FEC implementations.

The use of FEC limits assumes that the decoder's performance is fully characterized by $\mathrm{BER}_{\mathrm{pre}}$, and that different channels with the same $\mathrm{BER}_{\mathrm{pre}}$ will result in the same $\mathrm{BER}_{\mathrm{post}}$ using a given FEC code. Under some assumptions on independent bit errors (which can be achieved by interleaving the code bits), this assumption is justifiable, if the decoder is based on HDs. This is the case for HD-FEC, where the decoder is fed with bits modeled using a binary symmetric channel (BSC). The use of FEC limits, however, has not changed with the adoption of SD-FEC in optical communications, which has made the "SD-FEC limit" to become increasingly popular in the optical communications literature.

The application of SD-FEC in optical communications dates back to the pioneering experiments by Puc *et al.* in 1999 [19], who used a concatenation of a Reed–Solomon code and a convolutional code. Other early studies of SD include block turbo codes [20], [21] and low-density parity-check (LDPC) codes [22]–[24]. Another concatenated code suitable

Research supported by the Engineering and Physical Sciences Research Council (EPSRC) project UNLOC (EP/J017582/1), United Kingdom, and by the Swedish Research Council (VR) under grant no. 2012-5280. This work was presented in part at the 2015 Optical Fiber Communication Conference (OFC), Los Angeles, CA, Mar. 2015.

A. Alvarado, D. Lavery, R. Maher, and P. Bayvel are with the Optical Networks Group, Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, United Kingdom (email: alex.alvarado@ieee.org).

E. Agrell is with the Department of Signals and Systems, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.

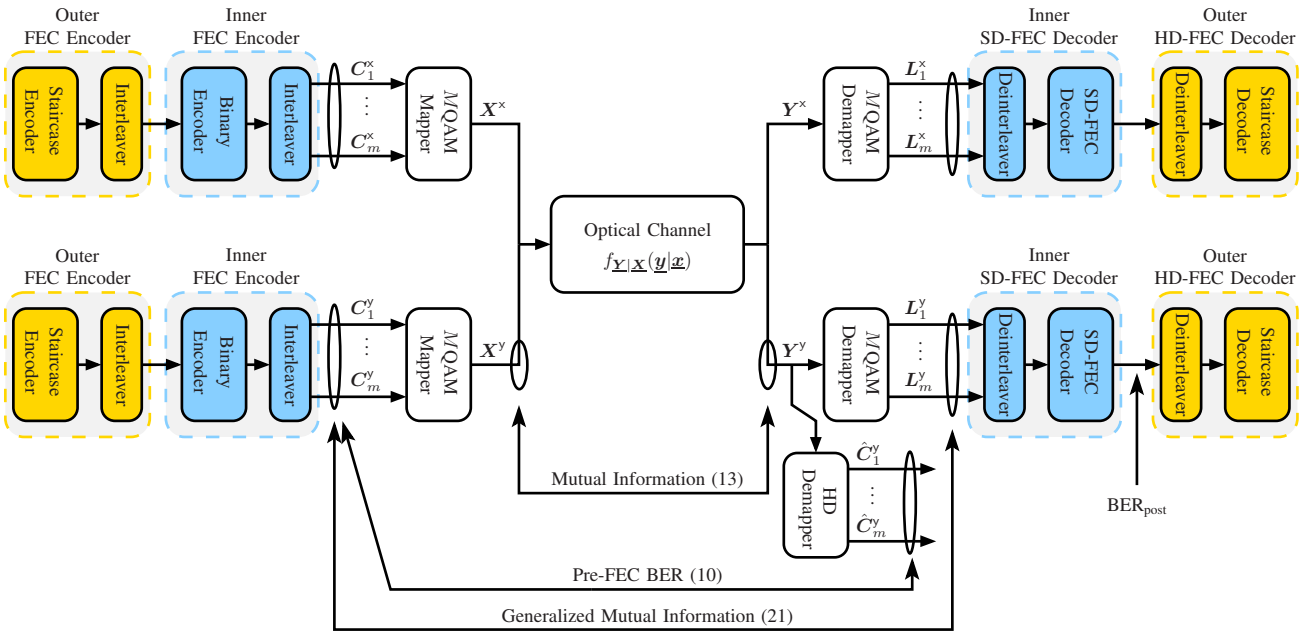


Fig. 1. Dual-polarization CM transceiver with SD-FEC under consideration. The transmitter for each polarization consists of two cascaded binary FEC encoders followed by an MQAM mapper. The receiver is a BW receiver: L-values are calculated by the demapper (ignoring the intersymbol and interpolarization interference), which is followed by an SD-FEC decoder and an HD-FEC decoder.

for SD decoding was defined for optical submarine systems by the ITU in the G.975.1 standard [25]. See [26], [27], and references therein for further details on SD-FEC in optical communications.

Tables and plots of BER_{post} vs. BER_{pre} were presented in, e.g., [21], [24], [25], under specific choices for the channel, modulation format, and symbol rate. Although this was not suggested when these tables and plots were originally published, the existence of such data has subsequently been adopted to avoid the need for including FEC in system simulations and experiments. This SD-FEC limit paradigm is nowadays very popular in optical communication system design. It has been used for example in the record experiments based on 2048 quadrature amplitude modulation (QAM) for single-core [28] and multi-core [29] fibers. It has, however, never been validated to which extent the function BER_{post} vs. BER_{pre}, determined for one set of system parameters (channel, modulation, symbol rate, etc.), accurately characterizes the same function with other parameters.

Another option to predict the post-FEC BER is to use the *mutual information* (MI) between the input and output of the discrete-time channel. This approach was suggested in [30]–[32] and applied to optical communications in [33] (see also [34]). In [33], it was shown that the MI is a better metric than the pre-FEC BER in predicting the post-FEC BER, which casts significant doubts on the SD-FEC limit paradigm.

This paper investigates the usage of the *generalized mutual information* (GMI) [1, Sec. 3], [2, Sec. 4.3] for the same purpose. The GMI, also known as the BICM capacity (or parallel decoding capacity), was introduced in an optical communications context in [7]. The performance of some LDPC codes with four-dimensional constellations over the additive white Gaussian noise (AWGN) channel was evaluated in terms of the GMI in [35]. With any given LDPC code, an apparent one-to-one mapping was observed between the GMI and the post-FEC BER, regardless of the constellation used. In this paper, which extends the conference version [36], we investigate this mapping further and show that the GMI is a very accurate post-FEC BER predictor, significantly more accurate than both the pre-FEC BER and the MI, under general conditions[1]. Consistent results were obtained for the nonlinear optical channel in both linear and nonlinear regimes, for the AWGN channel, for both LDPC codes and turbo codes, for a variety of modulation formats, and also validated by experiments.

This paper is organized as follows. In Sec. II, the system model is introduced and principles for FEC are reviewed. Sec. III introduces achievable rates, which are quantified by the MI and GMI. The post-FEC BER prediction is studied in Sec. IV. Conclusions are drawn in Sec. V.

II. Preliminaries

A. Channel and System Model

In this paper, we consider the CM transceiver shown in Fig. 1, which is the common for coherent optical communication systems. Data is transmitted in blocks of $2n_{\text{s}}$ symbols, where every block represents n_{s} time instants in each of the two polarizations. At the transmitter, an outer encoder is serially concatenated with an inner FEC encoder with code

[1]One of these conditions is that the binary code under consideration is *universal*, i.e., that its performance does not depend on the distribution of the soft information passed to the decoder, but only on the capacity of the channel [37, Sec. 9.5]. The universality property of LDPC codes for binary-input memoryless channels was initially discussed in [32], [38], later studied in, e.g., [39], [40], and recently for spatially-coupled LDPC codes in [41].

TABLE I
SUMMARY OF SYSTEM PARAMETERS USED IN WDM SIMULATION.

Parameter	Value
Fiber attenuation	0.2 dB/km
Dispersion parameter	17 ps/nm/km
Fiber nonlinear coefficient	1.2 $(\text{W km})^{-1}$
Span length	L km
PMD	0 ps/$\sqrt{\text{km}}$
Symbol rate	32 Gbaud
EDFA noise figure	3 dB
WDM channels	11
Channel separation	50 GHz
Pulse shape	RRC, 1% rolloff

rate R_c. The inner encoder generates code bits $\boldsymbol{C}_1^\mathsf{p}, \ldots, \boldsymbol{C}_m^\mathsf{p}$, where $\boldsymbol{C}_k^\mathsf{p} = [C_{k,1}^\mathsf{p}, C_{k,2}^\mathsf{p}, \ldots, C_{k,n_\text{s}}^\mathsf{p}]$, $k = 1, 2, \ldots, m$ is the bit position and $\mathsf{p} \in \{\mathsf{x}, \mathsf{y}\}$ indicates the polarization.[2] The code bits for each polarization are fed to a memoryless M-ary QAM (MQAM) mapper with $M = 2^m$ constellation points $\mathcal{X} \triangleq \{x_1, x_2, \ldots, x_M\}$. We consider Gray-mapped square QAM constellations with $M = 4, 16, 64, 256$ as well as (non-Gray) 8QAM from [42, Fig. 14 (a)].

The transmitted sequences of complex symbols $\boldsymbol{X}^\mathsf{p} = [X_1^\mathsf{p}, X_2^\mathsf{p}, \ldots, X_{n_\text{s}}^\mathsf{p}]$ with $X_l^\mathsf{p} \in \mathcal{X}$ is modulated using a root-raised-cosine (RRC) pulse with 1% rolloff. The symbols in the two polarizations are combined into the matrix

$$\underline{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{X}^\mathsf{x} \\ \boldsymbol{X}^\mathsf{y} \end{bmatrix} = \begin{bmatrix} X_1^\mathsf{x} & X_2^\mathsf{x} & \ldots & X_{n_\text{s}}^\mathsf{x} \\ X_1^\mathsf{y} & X_2^\mathsf{y} & \ldots & X_{n_\text{s}}^\mathsf{y} \end{bmatrix} \tag{1}$$

and sent through a nonlinear optical channel, whose parameters are summarized in Table I. We consider 11 dual-polarization wavelength-division multiplexed (WDM) channels of 32 Gbaud in a 50 GHz grid over a single span of single mode fiber (SMF) of length L with zero polarization mode dispersion (PMD). At the receiver, an erbium-doped fiber amplifier (EDFA) with an ideal noise figure of 3 dB (spontaneous emission factor $n_\text{sp} = 1$) is used. The digital signal processing (DSP) in the receiver includes electronic chromatic dispersion compensation (EDC) and matched filtering followed by ideal data-aided phase compensation. Data for the central channel is recorded and represented (for the two polarizations) by the received matrix $\underline{\boldsymbol{Y}}$ of size 2 by n_s, where $Y_l^\mathsf{p} \in \mathbb{C}$ for $l = 1, 2, \ldots, n_\text{s}$ and $\mathsf{p} \in \{\mathsf{x}, \mathsf{y}\}$.

The ideal phase compensation algorithm we consider compensates the nonlinear phase rotation of each received symbol by multiplying the received symbol by $\exp(-j\theta_i)$ with $i = 1, \ldots, M$, where θ_i is the average phase rotation experienced by the received symbols conditioned on the ith transmitted symbol. The average is taken over the entire codeword, whose length depends on the type of code and its code rate, as detailed in Sec. II-B.

As shown in Fig. 1, the optical channel is modeled by the channel law $f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{x}})$.[3] This discrete-time model encompasses all the transmitter DSP used after the MQAM mapper (i.e., pulse shaping and polarization multiplexing), the physical channel (the fiber and the EDFA), and the receiver DSP.

Even though some residual intersymbol interference usually remains after EDC and the received symbols are affected by interpolarization interference, these effects are typically ignored in current receivers, to reduce complexity. Hence, each symbol in $\underline{\boldsymbol{Y}}$ is decoded separately in both time and polarization. More specifically, for each $l = 1, \ldots, n_\text{s}$ and $\mathsf{p} \in \{\mathsf{x}, \mathsf{y}\}$, *soft* information on the code bits $C_{1,l}^\mathsf{p}, \ldots, C_{m,l}^\mathsf{p}$ is calculated in the form of *L-values*[4], also known as logarithmic likelihood ratios, as

$$L_{k,l}^\mathsf{p} \triangleq \log \frac{f_{Y_l^\mathsf{p}|C_{k,l}^\mathsf{p}}(y_l^\mathsf{p}|1)}{f_{Y_l^\mathsf{p}|C_{k,l}^\mathsf{p}}(y_l^\mathsf{p}|0)} \tag{2}$$

$$= L_{k,l}^{\mathsf{p},\text{apo}} - L_{k,l}^{\mathsf{p},\text{apri}} \tag{3}$$

where $k = 1, \ldots, m$ and

$$L_{k,l}^{\mathsf{p},\text{apo}} = \log \frac{P_{C_{k,l}^\mathsf{p}|Y_l^\mathsf{p}}(1|y_l^\mathsf{p})}{P_{C_{k,l}^\mathsf{p}|Y_l^\mathsf{p}}(0|y_l^\mathsf{p})}, \tag{4}$$

$$L_{k,l}^{\mathsf{p},\text{apri}} = \log \frac{P_{C_{k,l}^\mathsf{p}}(1)}{P_{C_{k,l}^\mathsf{p}}(0)} \tag{5}$$

are the *a posteriori* and *a priori* L-values, respectively.

A stationary channel model is assumed, and thus, the index l can be dropped. Furthermore, the performance in both polarizations is expected to be identical, so from now on, the notation $(\cdot)^\mathsf{p}$ is also dropped. Using this and the law of total probability in (2) gives

$$L_k = \log \frac{\sum_{x \in \mathcal{X}_k^1} P_{X|C_k}(x|1) f_{Y|X}(y|x)}{\sum_{x \in \mathcal{X}_k^0} P_{X|C_k}(x|0) f_{Y|X}(y|x)} \tag{6}$$

where $\mathcal{X}_k^b \subset \mathcal{X}$ is the set of constellation symbols labeled by a bit $b \in \mathbb{B} \triangleq \{0, 1\}$ at bit position $k \in \{1, \ldots, m\}$. The L-values calculated by the demapper are then passed to the SD-FEC decoder. The SD-FEC decoder makes a decision on the bits fed into the inner encoder. These bits are then used by the outer HD-FEC decoder, as shown in Fig. 1.

To alleviate the computational complexity of (6), the well-known max-log approximation [43]

$$L_k \approx \log \frac{\max_{x \in \mathcal{X}_k^1} P_{X|C_k}(x|1) f_{Y|X}(y|x)}{\max_{x \in \mathcal{X}_k^0} P_{X|C_k}(x|0) f_{Y|X}(y|x)} \tag{7}$$

is often used.

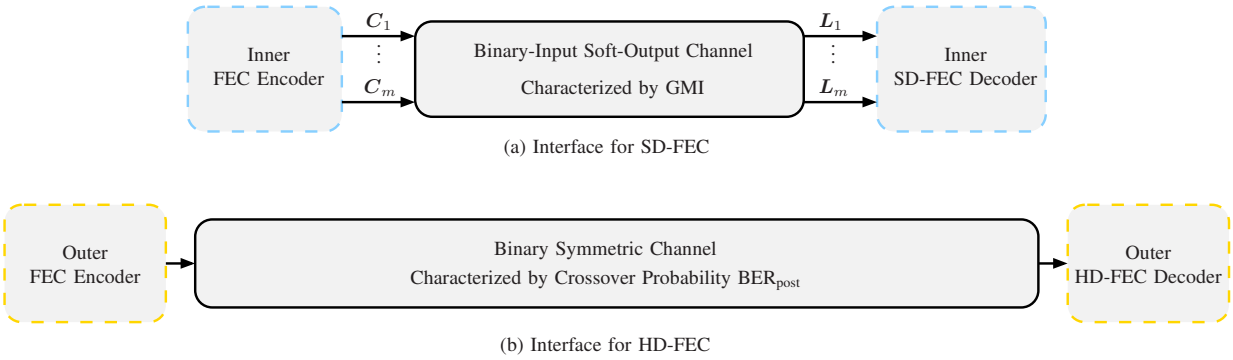
B. Pre-FEC BER

The lower branch of the receiver in Fig. 1 includes an HD demapper which makes an HD on the code bits. We assume that this HD demapper is the optimal memoryless HD demapper in the sense of minimizing the pre-FEC BER. This maximum *a posteriori* (MAP) decision rule is equivalent to

[2] Throughout this paper, boldface symbols denote random vectors.

[3] Throughout this paper, $f_A(a)$ denotes a probability density function (PDF) and $f_{A|B}(a|b)$ a conditional PDF. Similarly, $P_A(a) \triangleq \Pr\{A = a\}$ denotes a probability mass function (PMF) and $P_{A|B}(a|b) \triangleq \Pr\{A = a|B = b\}$ a conditional PMF.

[4] A sign operation on an L-value corresponds an HD. Its magnitude represents the reliability of the HD.



(a) Interface for SD-FEC

(b) Interface for HD-FEC

Fig. 2. Interface for (a) the inner SD-FEC and (b) the outer HD-FEC in Fig. 1 (for one polarization). The BISO channel characterized by the GMI and the BSC by its crossover probability given by the BER after SD-FEC decoding $\mathrm{BER}_{\mathrm{post}}$.

making an HD on the *a posteriori* L-values in (4): if $L_k^{\mathrm{apo}} \geq 0$ then $\hat{C}_k = 1$, and $\hat{C}_k = 0$ otherwise.[5] Formally,

$$\mathrm{BER}_{\mathrm{pre}} \triangleq \frac{1}{m}\sum_{k=1}^{m} \Pr\{\hat{C}_k \neq C_k\} \tag{8}$$

$$= \frac{1}{m}\sum_{k=1}^{m}\sum_{c\in\mathbb{B}} P_{C_k}(c)\Pr\{\hat{C}_k \neq c | C_k = c\} \tag{9}$$

$$= \frac{1}{m}\sum_{k=1}^{m}\sum_{c\in\mathbb{B}} P_{C_k}(c)\int_0^{\infty} f_{L_k^{\mathrm{apo}}|C_k}((-1)^c l|c)\,\mathrm{d}l. \tag{10}$$

The pre-FEC BER is a standard performance measure for uncoded systems. As discussed in Sec. II-D, pre-FEC BER is a good predictor of post-FEC BER for HD-FEC with ideal interleaving. We will show in Sec. IV that the pre-FEC BER is not necessarily a good predictor of post-FEC BER for SD-FEC.

C. SD-FEC: Turbo and LDPC Codes

We consider two families of binary SD-FEC: Turbo codes (TCs) and irregular repeat-accumulate LDPC codes. In both cases, a pseudo-random bit-level interleaver is assumed to be used prior to modulation (see Fig. 1). Without loss of generality, we assume this interleaver to be part of the inner FEC encoder.

The TCs we consider are formed as the parallel concatenation of two identical, eight-state, recursive and systematic convolutional encoders with code rate $1/2$. The generator polynomials are $(1, 11/15)_8$ [45] and the two encoders are separated by an internal random interleaver, giving an overall code rate $R_{\mathrm{c}} = 1/3$. Six additional code rates are obtained by cyclically puncturing parity bits using the patterns defined in [45] and [46], which leads to $R_{\mathrm{c}} \in \{1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 5/6\}$ and FEC overheads (OHs) of $\{200, 150, 100, 66.6, 50, 33.3, 20\}\%$. Each transmitted frame consists of $20,000$ information bits. The decoder is based on the max-log-MAP decoding algorithm with ten iterations. The extrinsic L-values exchanged during the iterations are scaled by 0.7 as suggested in [47].

The LDPC codes we consider are those proposed by the second generation satellite digital video broadcasting standard [48] with code rates $R_{\mathrm{c}} \in \{1/3, 2/5, 1/2, 3/5, 3/4, 9/10\}$. This leads to OHs of $\{200, 150, 100, 66.66, 33.3, 11.1\}\%$. Each transmitted frame consists of $64,800$ code bits. The decoder uses the message passing algorithm with 50 iterations and exact L-values.

What the SD-FEC encoder and decoder pair "sees" is a binary-input soft-output (BISO) channel. This is shown in Fig. 2 (a). This BISO channel is sometimes known in the literature as the BICM channel [49, Fig. 1] and it has been used to predict the decoder performance via probabilistic models of the L-values [2, Sec. 5.1]. In this paper, we are interested in finding a measure to characterize this BISO channel in order to predict the post-FEC BER across different channels.

D. HD-FEC: Staircase Codes

As shown in Fig. 1, the considered transceiver includes an outer encoder to reduce the BER after SD-FEC decoding to 10^{-15}. For both TCs and LDPC codes, we use the staircase code with 6.25% OH from [50, Table I]. For a BSC, this staircase code guarantees an output BER of 10^{-15} for a crossover probability of $4.7 \cdot 10^{-3}$. This corresponds to the HD-FEC limit paradigm, which is perfectly justifiable under the BSC assumptions.

To guarantee that the errors introduced by the inner SD-FEC decoder are independent within a frame, we include a bit-level interleaver (see Fig. 1). Under these assumptions, what the HD-FEC encoder and decoder pair "sees" is a BSC with crossover probability given by the BER after SD-FEC decoding ($\mathrm{BER}_{\mathrm{post}}$). Therefore, the BER after HD-FEC decoding can be assumed to be 10^{-15} for $\mathrm{BER}_{\mathrm{post}} = 4.7 \cdot 10^{-3}$. This is shown in Fig. 2 (b). From now on, we therefore assume the existence of the interleaver and staircase code, and thus, without loss of generality, we focus on a target BER after SD-FEC decoding of $\mathrm{BER}_{\mathrm{post}} = 4.7 \cdot 10^{-3}$.

[5] This decision rule is slightly better than the standard demapper based on HDs on the symbols followed by a symbol-to-bit mapper (inverting the bit-to-symbol mapping used at the transmitter). However, the differences are noticeable only at very high pre-FEC BER [44, Sec. V].

III. Achievable Rates

Achievable rates indicate the number of bits per symbol that can be reliably transmitted through the channel. In this section we review achievable rates for channels with memory, for optimal decoders, and for BW decoders. These achievable rates will be used in Sec. IV to predict the post-FEC BER.

A. Channels with Memory

A *coding scheme* consists of a codebook, an encoder, and a decoder. The codebook is the set of codewords that can be transmitted through the channel, where each codeword is a sequence of symbols. The encoder is a one-to-one mapping between the information sequences and codewords. The decoder is a deterministic rule that maps the noisy channel observations onto an information sequence.

A code rate, in bits per (single-polarization) symbol, is said to be *achievable* at a given block length n_s and for a given average error probability ε if there exists a coding scheme whose average error probability is below ε. Under certain assumptions on information stability [51, Sec. I], and for any stationary random process $\{X_l\}$ with joint PDF $f_{\underline{\boldsymbol{X}}}$, an achievable rate for channels with memory (i.e., where symbols are correlated in time and across polarizations) is given by

$$R^{\mathrm{mem}} = \lim_{n_\mathrm{s}\to\infty} \frac{1}{2n_\mathrm{s}} I(\underline{\boldsymbol{X}};\underline{\boldsymbol{Y}}) \tag{11}$$

where $I(\underline{\boldsymbol{X}};\underline{\boldsymbol{Y}})$ is the mutual information defined as

$$I(\underline{\boldsymbol{X}};\underline{\boldsymbol{Y}}) \triangleq \mathbb{E}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}\left[\log_2 \frac{f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}}}(\underline{\boldsymbol{Y}}|\underline{\boldsymbol{X}})}{f_{\underline{\boldsymbol{Y}}}(\underline{\boldsymbol{Y}})}\right] \tag{12}$$

and where $\mathbb{E}_{\underline{\boldsymbol{X}},\underline{\boldsymbol{Y}}}$ denotes the expectation with respect to both $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$. The *channel capacity* is the largest achievable rate for which a coding scheme with vanishing error probability exists, in the limit of large block length.

B. Memoryless Receivers

Although the discrete-time optical channel in Sec. II-A suffers from intersymbol and interpolarization interference, the standard receiver considered in this paper ignores these effects. In particular, each polarization is considered independently (see Fig. 1), and the soft information on the coded bits is calculated ignoring correlation between symbols in time (see (2)). To model these assumptions made by the receiver, the channel is modeled by a conditional PDF $f_{Y|X}(Y|X)$. Therefore, from now on, and without loss of generality, only one polarization is considered. Furthermore, we assume the symbols are independent random variables drawn from a distribution f_X.

An achievable rate for transceivers that ignore intersymbol and interpolarization interference is

$$I(X;Y) = \mathbb{E}_{X,Y}\left[\log_2 \frac{f_{Y|X}(Y|X)}{f_Y(Y)}\right]. \tag{13}$$

where $I(X;Y)$ is the unidimensional version of the MI in (12). As expected, $R^{\mathrm{mem}} \geq I(X;Y)$ [52, Sec. III-F] and thus, $I(X;Y)$ is a (possibly loose) lower bound on the capacity of the channel with intersymbol and interpolarization interference.

Let $\mathcal{C}$ the binary codebook used for transmission and $\underline{\boldsymbol{c}}$ denote the transmitted codewords as

$$\underline{\boldsymbol{c}} = \begin{bmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,n_\mathrm{s}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \dots & c_{m,n_\mathrm{s}} \end{bmatrix}. \tag{14}$$

Furthermore, let $\boldsymbol{B} = [B_1,\dots,B_m]$ be a random vector representing the transmitted bits $[c_{1,l},\dots,c_{m,l}]$ at any time instant l, which are mapped to the corresponding symbol $X_l \in \mathcal{X}$ with $l = 1,2,\dots,n_\mathrm{s}$. Assuming a memoryless channel and equally likely codewords, the optimal maximum-likelihood (ML) receiver chooses the transmitted codeword based on an observed sequence $[y_1,\dots,y_n]$ according to the rule

$$\underline{\boldsymbol{c}}^{\mathrm{ml}} \triangleq \operatorname*{argmax}_{\underline{\boldsymbol{c}}\in\mathcal{C}} \sum_{l=1}^{n_\mathrm{s}} \log f_{Y|\boldsymbol{B}}(y_l|c_{1,l},\dots,c_{m,l}). \tag{15}$$

Shannon's channel coding theorem states that reliable transmission with the ML decoder in (15) is possible at arbitrarily low error probability if the combined rate of the binary encoder and mapper (in information bit/symbol) is below $I(X;Y)$, i.e., if $R_\mathrm{c} m \leq I(X;Y)$.

For a discrete constellation $\mathcal{X}$, the MI in (13) can be expressed as

$$I(X;Y) = \sum_{x\in\mathcal{X}} P_X(x) \int_{\mathbb{C}} f_{Y|X}(y|x) \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)}\,\mathrm{d}y. \tag{16}$$

A Monte-Carlo estimate thereof is

$$I(X;Y) \approx \frac{1}{n_\mathrm{s}} \sum_{x\in\mathcal{X}} P_X(x) \sum_{l=1}^{n_\mathrm{s}} \log_2 \frac{f_{Y|X}(t^{(l)}|x)}{f_Y(t^{(l)})} \tag{17}$$

where $t^{(l)}$ with $l = 1,2,\dots,n_\mathrm{s}$ are independent and identically distributed (i.i.d.) random variables distributed according to the channel law $f_{Y|X}(y|x)$.

C. BW Receivers

As shown in Fig. 1, the BW decoder considered in this paper splits the decoding process. First, L-values are calculated, and then, a binary SD decoder is used. More precisely, the BW decoder rule is

$$\underline{\boldsymbol{c}}^{\mathrm{bw}} \triangleq \operatorname*{argmax}_{\underline{\boldsymbol{c}}\in\mathcal{C}} \sum_{l=1}^{n_\mathrm{s}} \log \prod_{k=1}^{m} f_{Y|B_k}(y_l|c_{k,l}). \tag{18}$$

The BW decoding rule in (18) is not the same as the ML rule in (15) and the MI is in general not an achievable rate with a BW decoder.[6]

The channel under consideration is a symbol-wise channel defined by the channel transition probability $f_{Y|X}(y|x)$. The

[6] An exception is the trivial case of Gray-mapped 4QAM (i.e., quadrature phase-shift keying, QPSK) with noise added in each quadrature independently. In this case, the detection can be decomposed into two binary phase-shift keying constellations, and thus, ML and BW decoders are identical.

BW decoder can therefore be cast into the framework of a mismatched decoder by considering a symbol-wise metric

$$q(\boldsymbol{b}, y) \triangleq \prod_{k=1}^{m} f_{Y|B_k}(y|b_k). \tag{19}$$

Using this mismatched decoding formulation, the BW rule in (18) can be expressed as

$$\underline{\boldsymbol{c}}^{\mathrm{bw}} = \underset{\underline{\boldsymbol{c}} \in \mathcal{C}}{\operatorname{argmax}} \sum_{l=1}^{n_\mathrm{s}} \log q(\boldsymbol{b}_l, y_l) \tag{20}$$

where with a slight abuse of notation we use $\boldsymbol{b}_l = [c_{1,l}, \ldots, c_{m,l}]^T$. In this context, the ML decoder in (15) can be seen as a mismatched decoder with a metric $q(\boldsymbol{b}_l, y_l) = f_{Y|\boldsymbol{B}}(y_l|\boldsymbol{b}_l) = f_{Y|X}(y_l|x_l)$ which is "matched" to the symbol-wise channel. Using this interpretation, the BW decoder uses metrics matched to the bits $f_{Y|B_k}(y|b_k)$, but not matched to the actual channel, and thus, its name.

An achievable rate for a BW decoder is the GMI, which represents a bound on the number of bits per symbol that can be reliably transmitted through the channel. The GMI is defined as [53, eq. (59)–(60)] [2, (4.34)–(4.35)]

$$\mathrm{GMI} \triangleq \max_{s \geq 0} \mathbb{E}_{\boldsymbol{B},Y} \left[\log_2 \frac{q(\boldsymbol{B}, Y)^s}{\sum_{\boldsymbol{b} \in \mathbb{B}^m} P_{\boldsymbol{B}}(\boldsymbol{b}) q(\boldsymbol{b}, Y)^s} \right]. \tag{21}$$

For the BW metric in (19) and assuming independent bits $B_1, \ldots, B_m$, the GMI in (21) can be expressed as

$$\mathrm{GMI} = \max_{s \geq 0} \sum_{k=1}^{m} \mathbb{E}_{B_k,Y} \left[\log_2 \frac{f_{Y|B_k}(Y|B_k)^s}{\sum_{b \in \mathbb{B}} P_{B_k}(b) f_{Y|B_k}(Y|b)^s} \right] \tag{22}$$

$$= \sum_{k=1}^{m} \mathbb{E}_{B_k,Y} \left[\log_2 \frac{f_{Y|B_k}(Y|B_k)}{\sum_{b \in \mathbb{B}} P_{B_k}(b) f_{Y|B_k}(Y|b)} \right] \tag{23}$$

$$= \sum_{k=1}^{m} I(B_k; Y) \tag{24}$$

where (22) follows from [2, Theorem 4.11] and (23) from [2, Corollary 4.12] (obtained with $s = 1$). The expression in (24) follows from the definition of MI in (13).

In general, $I(X;Y) \geq$ GMI [2, Theorem 4.24][7], where the rate penalty $I(X;Y) -$ GMI can be understood as the penalty caused by the use of a suboptimal (BW) decoder. This rate penalty, however, is known to be small for Gray-labeled constellations [4, Fig. 4], [54], [55], [56, Sec. IV].

The GMI has not been proven to be the largest achievable rate for the receiver in Fig. 1. For example, a different achievable rate—the so-called LM rate—has been recently studied in [57, Part I]. Moreover, in the case where unequally likely constellation points are allowed, a new achievable rate has been recently derived in [58, Theorem 1]. Finding the largest achievable rate with a BW decoder remains an open research problem. Despite this cautionary statement, the GMI is known to predict well the performance of CM transceivers based on capacity-approaching SD-FEC decoders. This will be shown in Sec. IV.

[7]The condition of i.i.d. bits in [2, Theorem 4.24] is not necessary—only independence is needed.

When the L-values are calculated using (6), $I(B_k;Y) = I(B_k;L_k)$ [2, Theorem 4.21], and thus, the GMI in (24) becomes

$$\mathrm{GMI} = \sum_{k=1}^{m} I(B_k; L_k) \tag{25}$$

i.e., the GMI is a sum of *bit-wise* MIs between code bits and L-values. The equality in (25) does not hold, however, if the L-values were calculated using the max-log approximation (7), or more generally, if the L-values were calculated using any other approximation. For example, when max-log L-values are considered, it is possible to show that there is a loss in achievable rates. Under certain conditions, this loss can be recovered by adapting the max-log L-values, as shown in [59]–[61].

Regardless of the L-value calculation, the GMI in (22) can be estimated via Monte-Carlo integration as [2, Theorem 4.20]

$$\mathrm{GMI} \approx \sum_{k=1}^{m} H_\mathrm{b}(P_{B_k}(0)) - \frac{1}{n_\mathrm{s}} \min_{s \geq 0} \sum_{k=1}^{m} \sum_{b \in \mathbb{B}} P_{B_k}(b) \sum_{n=1}^{n_\mathrm{s}} \log_2 \left(1 + e^{s(-1)^b \lambda_{k,b}^{(n)}} \right) \tag{26}$$

where $\lambda_{k,b}^{(n)}$, $n = 1, 2, \ldots, n_\mathrm{s}$ are i.i.d. random variables distributed according to the PDF of the L-values $f_{L_k|B_k}(\lambda|b)$ and $H_\mathrm{b}(p) \triangleq -p \log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy function. The minimization over s in (26) can be easily approximated (numerically) using the concavity of the GMI on s [2, eq. (4.81)].

We emphasize here that the expression in (26) is valid for any symbol-wise metric in the form of (19), i.e., for any L-value L_k that ignores the dependency between the bits in the symbol. In particular, when the L-values are calculated exactly using (6), the GMI can be estimated using (26) and $s = 1$, which follows from [2, Theorem 4.20].

D. Memoryless and BW Receivers for the AWGN Channel

Often, if not always, CM transceivers in optical communication systems assume that the discrete-time channel, including transmitter- and receiver-side DSP, is a memoryless AWGN channel $Y = X + Z$, where Z is a complex, zero-mean, circularly symmetric Gaussian random variable with total variance $\mathbb{E}[|Z|]^2$. This assumption might be suboptimal, but in the absence of a better (non-Gaussian) model with memory, the memoryless AWGN channel assumption is reasonable. In this subsection, we specialize the MI and GMI estimators in (17) and (26) to the AWGN channel and equally likely input bits (and therefore, equally likely symbols in $\mathcal{X}$).

For the AWGN channel and a uniform input distribution, the MI in (16) can be estimated using (17) as

$$I(X;Y) \approx \log_2(M) - \frac{1}{M n_\mathrm{s}} \sum_{i=1}^{M} \sum_{l=1}^{n_\mathrm{s}} \log_2 f_{i,l}, \tag{27}$$

where

$$f_{i,l} \triangleq \sum_{j=1}^{M} \exp\left(-\rho(2\Re\{(x_i - x_j)^* z^{(l)}\} + |z^{(l)}|^2) \right), \tag{28}$$

the signal-to-noise ratio (SNR) ρ is defined as $\rho \triangleq \mathbb{E}_X[|X|^2]/\mathbb{E}[|Z|]^2$, and $z^{(l)}$ with $l = 1, 2, \ldots, n_\mathrm{s}$ are n_s independent realizations of the Gaussian random variable Z.

L-values may be calculated either exactly or using the max-log approximation. In the first case, the exact L-values in (6) are calculated as

$$L_k = \log \frac{\sum_{x \in \mathcal{X}_k^1} \exp(-\rho|y-x|^2)}{\sum_{x \in \mathcal{X}_k^0} \exp(-\rho|y-x|^2)} \tag{29}$$

where we used the uniform input symbol distribution assumption. For given sequences of mn_s transmitted bits $c_{k,l}$ and mn_s L-values $\lambda_{k,l}$ computed via (29), for $k = 1, \ldots, m$ and $l = 1, \ldots, n_\mathrm{s}$, the GMI in (26) can be estimated as

$$\mathrm{GMI} \approx m - \frac{1}{n_\mathrm{s}} \sum_{k=1}^{m} \sum_{l=1}^{n_\mathrm{s}} \log_2\left(1 + e^{(-1)^{c_{k,l}} \lambda_{k,l}}\right). \tag{30}$$

In the second case, the max-log L-values in (7) are calculated as

$$L_k \approx \rho \left(\min_{x \in \mathcal{X}_k^0} |y-x|^2 - \min_{x \in \mathcal{X}_k^1} |y-x|^2 \right). \tag{31}$$

For given sequences of transmitted bits $c_{k,l}$ and max-log L-values $\lambda_{k,l}$ computed via (31), the GMI can be estimated using (26) as

$$\mathrm{GMI} \approx m - \frac{1}{n_\mathrm{s}} \min_{s \geq 0} \sum_{k=1}^{m} \sum_{l=1}^{n_\mathrm{s}} \log_2\left(1 + e^{s(-1)^{c_{k,l}} \lambda_{k,l}}\right). \tag{32}$$

To calculate the GMI, (30) and (32) should be used for exact and max-log L-values, respectively. Using (30) for max-log L-values results in a rate lower than the true one, i.e., the minimization over s in (32) is a mandatory step for approximated L-values. Furthermore, the distribution of the LLRs for nonbinary modulation is in general unknown. However, in the case of max-log LLRs and Gray-labeled squared QAM constellations, the distribution is known to be a sum of piecewise Gaussian functions [62].

IV. Post-FEC BER Prediction

In this section, we study the robustness of three different metrics to predict the post-FEC BER of SD-FEC: the pre-FEC BER, the MI, and the GMI. The aim is to find a robust and easy-to-measure metric that can be used to predict the post-FEC BER of a given encoder and decoder pair across different channels. Results for the AWGN channel are shown first, followed by results for the nonlinear optical channel. All the results were obtained by transmitting at least 500 codewords and by counting at least 100 bits in error after decoding.

A. AWGN Channel

To study the post-FEC BER prediction *across* different BISO channels (see Fig. 2 (a)), we consider the TCs defined in Sec. II-C and four modulation formats: M-QAM constellations with $M = 4, 8, 64, 256$. For $M = 4, 64$, the SD decoder uses exact L-values and for $M = 8, 256$, max-log L-values[8].

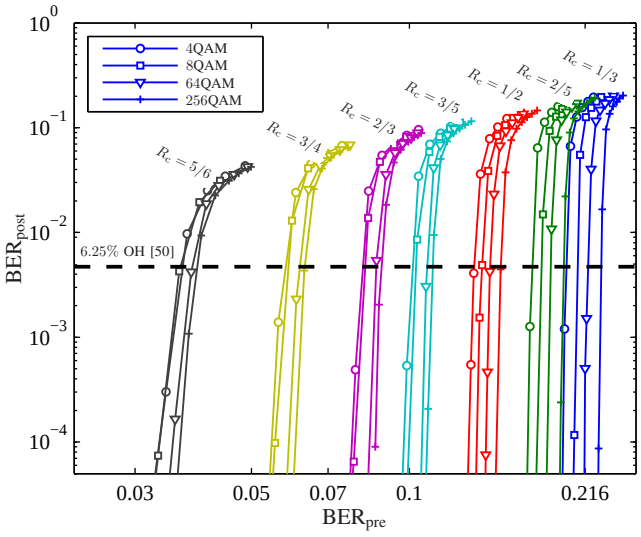
[8]For $M = 256$, the use of max-log L-values is very relevant in practice as the calculation in (29) is greatly simplified.

In Fig. 3 (a), the post-FEC BER is shown as a function of $\mathrm{BER_{pre}}$ for the 28 cases. Ideally, all the lines for the same rate (same color) should fall on top of one another, indicating that measuring $\mathrm{BER_{pre}}$ is enough to predict $\mathrm{BER_{post}}$ when the BISO channel (in this case, the modulation format) changes. The results in this figure show that this is not the case, especially for low and medium code rates. The pre-FEC BER therefore fails to predict the performance of the SD-FEC decoder across different BISO channels.

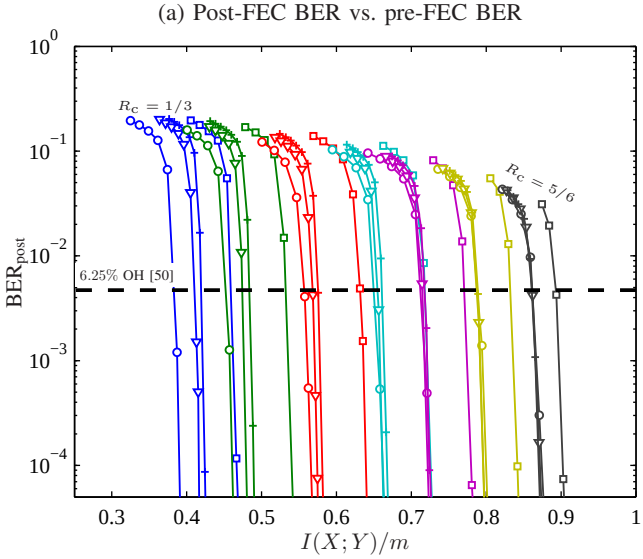
To estimate the inaccuracy of the SD-FEC limit paradigm, consider the results for 4QAM and $R_\mathrm{c} = 1/3$ shown in Fig. 3 (a). For a target post-FEC BER of $\mathrm{BER_{post}} = 4.7 \cdot 10^{-3}$, the required pre-FEC BER is $\mathrm{BER_{pre}} \approx 0.2$. By using the SD-FEC limit paradigm, we can conclude that to guarantee the same for post-FEC BER for 256QAM, the same pre-FEC BER can be assumed ($\mathrm{BER_{pre}} \approx 0.2$). This is clearly not the case, as for 256QAM and $R_\mathrm{c} = 1/3$, the pre-FEC BER can be higher ($\mathrm{BER_{pre}} \approx 0.23$). An alternative interpretation of this is that the results in Fig. 3 (a) show that for $\mathrm{BER_{pre}} \approx 0.2$ and 256QAM, the code rate can be increased to $R_\mathrm{c} = 2/5$. This shows that the use of the the SD-FEC limit paradigm in this scenario leads to an underestimation of the spectral efficiency of 20%. Very similar conclusions can be in fact drawn for the LDPC codes shown in [35, Fig. 4]. We also conjecture that the use of the SD-FEC limit paradigm in the record results reported in [28], [29] (where a pre-FEC BER threshold obtained for 4QAM was used for 2048QAM) are in fact incorrect and even higher spectral efficiencies can be obtained. In this case, however, the considered code rates are relatively high, and thus, we expect the underestimation to be below 5%.

The results in Fig. 3 (a) show the variations on the required pre-FEC BER to guarantee a given post-FEC BER across different modulation formats. While for low code rates these variations could lead to errors of up to 20% in spectral efficiencies, the errors decrease as the code rate increases. This partially suggests that the pre-FEC BER is a relatively good metric for high code rates, however, we have no theoretical justification for the use of $\mathrm{BER_{pre}}$ to predict the performance of a SD-FEC. Furthermore, we believe that having a metric that works for all code rates is important. Considering only high code rates—as is usually done in the optical community—is an artificial constraint that reduces flexibility in the design, as pointed out in [63, Sec. II-B].

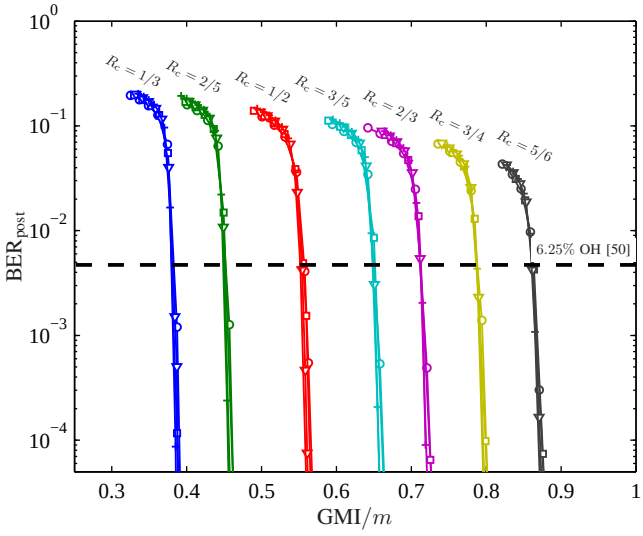
An intuitive explanation for the results in Fig. 3 (a) is that the SD-FEC in Fig. 1 does not operate on bits, and thus, a metric that is based on bits (i.e., the pre-FEC BER) cannot be used to predict the performance of the decoder. To clarify this, we compare 8QAM and 64QAM for $R_\mathrm{c} = 1/3$ and a target $\mathrm{BER_{pre}} \approx 0.216$. Exact L-value calculations are considered in both cases. From Fig. 3 (a) we see that $\mathrm{BER_{post}} \approx 5 \cdot 10^{-4}$ for 64QAM. For 8QAM, this value is $\mathrm{BER_{post}} \approx 5 \cdot 10^{-2}$, which is slightly lower than the one shown in Fig. 3 (a) for max-log



(a) Post-FEC BER vs. pre-FEC BER



(b) Post-FEC BER vs. normalized MI



(c) Post-FEC BER vs. normalized GMI

Fig. 3. Post-FEC BER for TCs with $R_c \in \{1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 5/6\}$ (colors) and different modulation formats (markers): 4QAM, 8QAM, 64QAM, and 256QAM. The post-FEC BER is shown versus (a) pre-FEC BER, (b) normalized MI, and (c) normalized GMI. The L-values for 8QAM and for 256QAM are calculated using the max-log approximation.

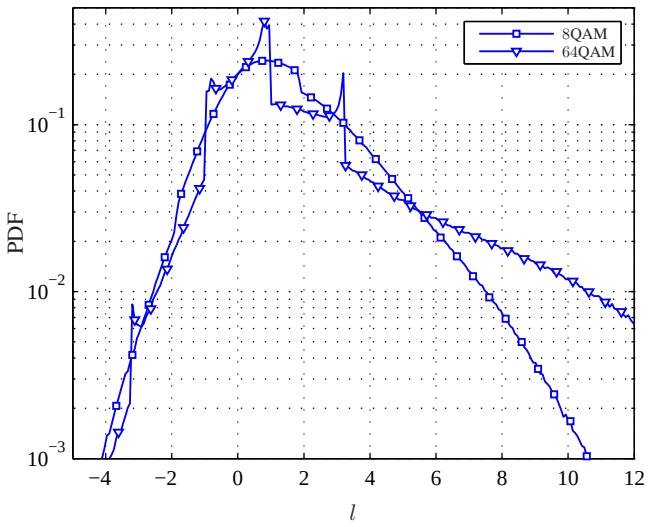


Fig. 4. Conditional PDF of the L-values $f_{L|B}(l|1)$ in (33) for 8QAM and 64QAM. In both cases, the L-values are calculated using (29). The PDFs of the L-values give identical pre-FEC BER but different post-FEC BER.

L-values. In Fig. 4 we show the PDF[9]

$$f_{L|B}(l|b) = \frac{1}{2m}\sum_{k=1}^{m} f_{L_k|B_k}(l|b) + f_{L_k|B_k}(-l|1-b). \quad (33)$$

The PDF in (33) corresponds to the conditional PDF of "symmetrized" and "mixed" L-values. For exact L-values, this PDF has been recently shown in [64, Sec. V] to fully determine the GMI (via $\mathrm{GMI} = mI(B;L)$). Under the uniform bit probability assumption, the pre-FEC BER in (10) can be expressed as

$$\mathrm{BER_{pre}} = \frac{1}{2m}\sum_{k=1}^{m}\int_{-\infty}^{0} \left(f_{L_k|B_k}(-l|0) + f_{L_k|B_k}(l|1)\right) \mathrm{d}l \quad (34)$$

and thus, it is clear that the pre-FEC BER can be calculated by

$$\mathrm{BER_{pre}} = \int_{-\infty}^{0} f_{L|B}(l|1)\,\mathrm{d}l \quad (35)$$

where $f_{L|B}(l|b)$ is given by (33).

While both PDFs $f_{L|B}(l|1)$ in Fig. 4 give the same pre-FEC BER ($\mathrm{BER_{pre}} \approx 0.216$), the post-FEC BER for 64QAM is much lower than the one for 8QAM. This can be explained by the different shapes of the PDFs in Fig. 4. In particular, the slow-decaying right tail of the PDF of 64QAM shows that some L-values with high reliability (i.e., high magnitude) will be observed, which the iterative SD-FEC decoder can exploit.

To study the dependency of the post-FEC BER and GMI on the distribution of the LLRs, we now compare 4QAM and 64QAM with $R_c = 1/2$. In particular, we consider the two SNR values that give $\mathrm{BER_{post}} \approx 4.7 \cdot 10^{-3}$ for the two modulation formats. Fig. 5 shows the PDF of the L-values for these two cases. Although the PDFs in Fig. 5 are clearly different, in both cases the normalized GMIs are identical

[9]Estimated via histograms.

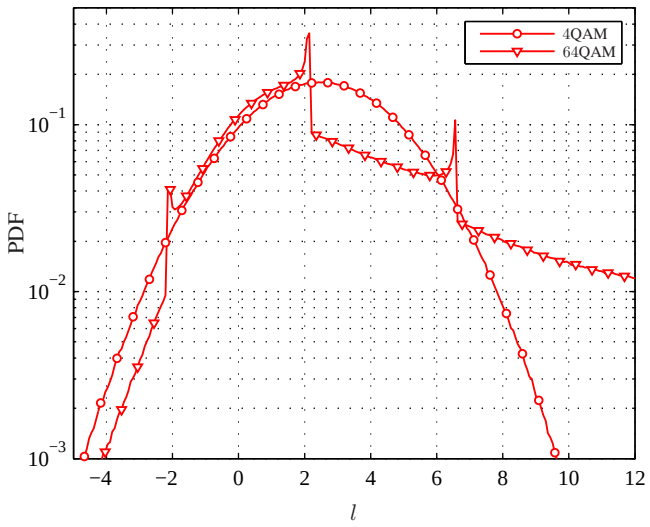
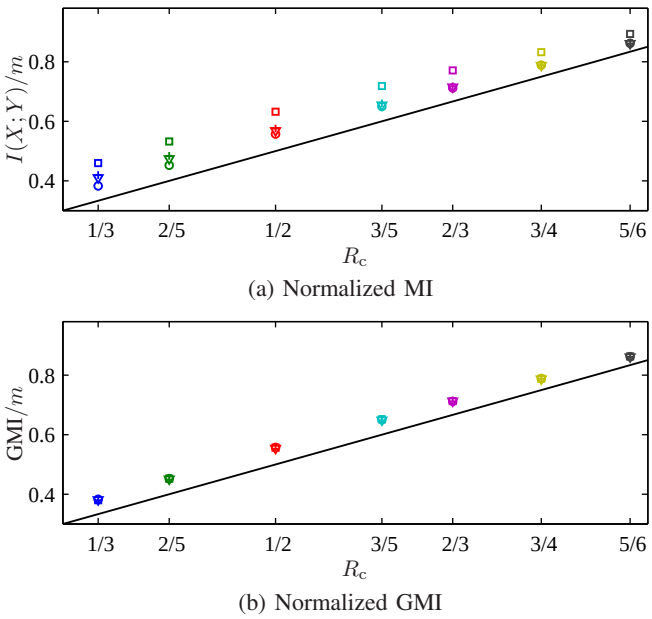


Fig. 5. Conditional PDF of the L-values $f_{L|B}(l|1)$ in (33) for 4QAM and 64QAM. In both cases, the L-values are calculated using (29). The PDFs of the L-values give identical normalized GMI and also identical post-FEC BER.



(a) Normalized MI

(b) Normalized GMI

Fig. 6. Required values for the different metrics to give $\text{BER}_{\text{post}} = 4.7 \cdot 10^{-3}$ as a function of the code rate for the same cases as in Fig. 3: (a) normalized MI and (b) normalized GMI. The curves $I(X;Y) = mR_c$ and $\text{GMI} = mR_c$ are shown in (a) and (b), resp.

(GMI/$m \approx 0.55$). This shows that the shape of the PDF of the L-values is not relevant for the post-FEC BER prediction. What is important is the normalized GMI obtained from those L-values.

Using BER_{pre} to predict the performance of SD-FEC decoders has no information-theoretic justification. To remedy this, one could consider the symbol-wise MI $I(X;Y)$ (see Fig. 1) as a metric to better predict BER_{post}. The values of BER_{post} as a function of the normalized MI $I(X;Y)/m$ are shown in Fig. 3 (b). In this case too, the prediction does not work well across all rates. In particular, although for square QAM constellations ($M = 4, 64, 256$) the MI seems to work well for high code rates (as previously reported in [35, Sec. III]), this is not the case if 8QAM is considered. The MI then appears to be less reliable to predict BER_{post} than the pre-FEC BER.

One intuitive explanation for the results in Fig. 3 (b) is that the MI is an achievable rate for the *optimum* receiver in (15), but not for the (suboptimal) receiver in Fig. 1 (see (18)). Another explanation is related to the performance dependence of BER_{post} on the binary labeling of the constellation. It is nowadays well understood that for the receiver in Fig. 1, the performance of the SD-FEC decoder depends on the binary labeling; Gray (or quasi-Gray) labelings are known to be among the best. On the other hand, the MI does not depend on the binary labeling but only on the constellation. Thus, it is not surprising that a labeling-independent metric fails at predicting the labeling-dependent BER_{post}.

The third and last metric we consider to predict BER_{post} is the GMI. The rationale behind this is that an SD-FEC decoder is fed with L-values, and thus, the GMI (see (25)) is an intuitively reasonable metric. The values of BER_{post} as a function of the normalized GMI are shown in Fig. 3 (c). These results show that for a given code rate, changing the constellation does not affect the post-FEC BER prediction based on the GMI. More importantly, and unlike for BER_{post} and MI, the prediction based on the GMI appears to work across *all* code rates. These results in fact show that the considered TCs appear to be universal (with respect to the GMI), which, to the best of our knowledge, has never been shown in the literature.

Fig. 6 shows the values of MI and GMI needed for each configuration in Fig. 3 to reach a post-FEC BER of $\text{BER}_{\text{post}} = 4.7 \cdot 10^{-3}$.[10] These values are obtained by finding the crossing points of the curves in Fig. 3 and the horizontal dashed lines. Fig. 6 also shows the relationships $I(X;Y) = mR_c$ and $\text{GMI} = mR_c$, where the vertical difference between the markers and the solid lines represent the rate penalty for these codes. The results in Fig. 6 clearly show the excellent prediction based on GMI and how MI does not work well across different modulation formats.

We showed before that the pre-FEC BER can lead to an erroneous estimate of the spectral efficiency, which is particularly noticeable for low code rates. A similar problem occurs if the normalized MI in Fig. 6 (a) is used to predict post-FEC BER. For example, the results in Fig. 6 (a) show that post-FEC BER of $\text{BER}_{\text{post}} = 4.7 \cdot 10^{-3}$ can be achieved with 4QAM and $R_c = 2/3$ when the normalized MI is approximately 0.71 (see also Fig. 3 (b)). One might be tempted to then conclude that, for the same MI, the same post-FEC BER can be achieved with 8QAM and $R_c = 2/3$. The results in Fig. 6 (a) show that this is in fact not possible, and a (lower) code rate of $R_c = 3/5$ is needed. In other words, the use of a "MI threshold paradigm" could lead to an overestimation (in this case by 11%) of the true spectral efficiency. This is not the case for

[10] Similar results could be presented in terms of pre-FEC BER. To have a fair comparison in terms of rates, however, one would need to convert the pre-FEC BER into SNR, and then map that SNR onto MI (or GMI), giving exactly what is shown in Fig. 3.

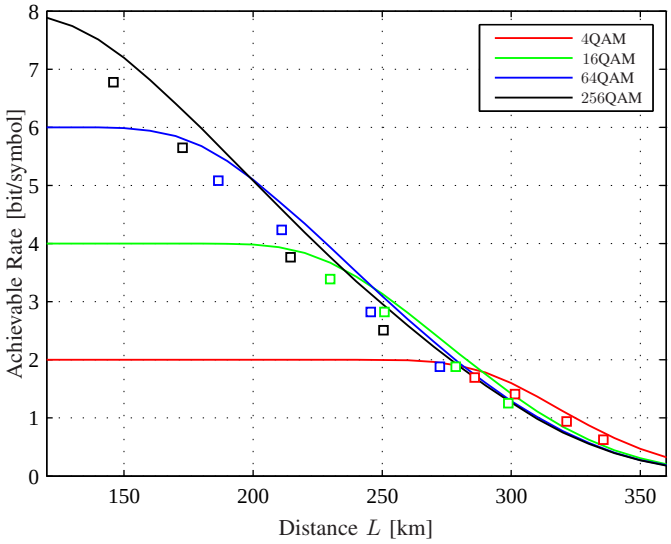


Fig. 7. Achievable rates (per polarization) versus span length: GMI (solid lines) and LDPC codes with $R_c \in \{1/3, 1/2, 3/4, 9/10\}$ (markers).

the GMI (see Fig. 6 (b)), where all markers for the same code fall on top of one another.

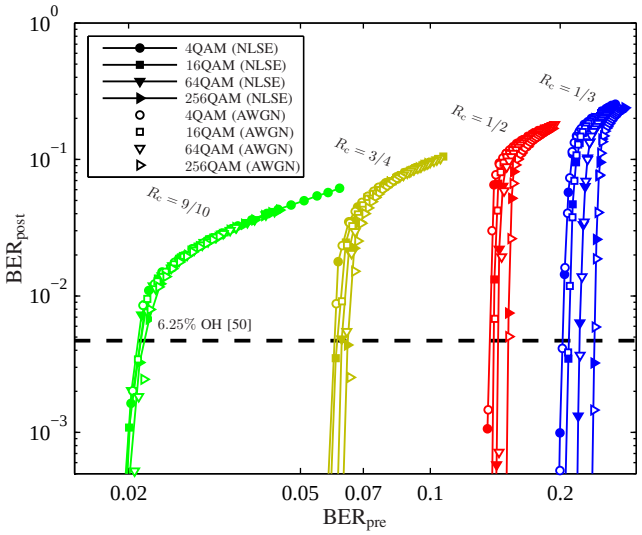
B. Optical Channel—Simulations

Dual-polarization transmission over the nonlinear optical channel specified in Sec. II-A was simulated using the coupled polarization nonlinear Schrödinger equation (NLSE) [65, eq. (6)]. This enabled the consideration of an idealized transmission link with zero polarization mode dispersion. The simulations were carried out via the split-step Fourier method with a step size of 100 m and an oversampling factor of 4 samples/symbol.

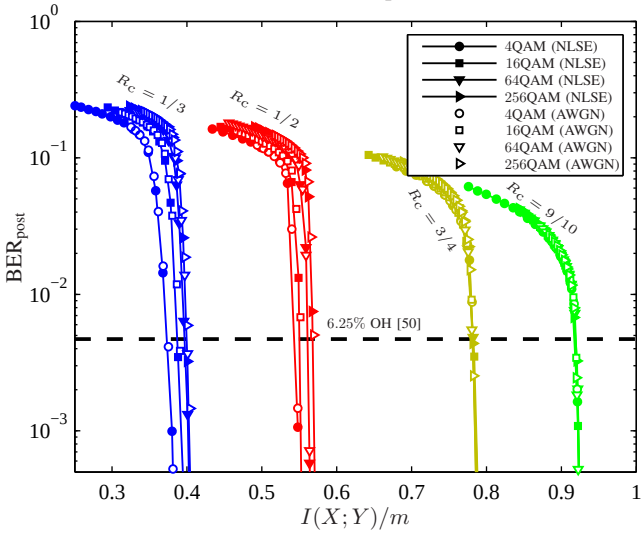
Fig. 7 shows the GMI (per polarization) as a function of the span length, for MQAM constellations with $M = 4, 16, 64, 256$. For each distance and M, we used the launch power that gave the highest GMI. In this figure, we also show the distance required by the LDPC codes in Sec. II-A to give $\mathrm{BER_{post}} = 4.7 \cdot 10^{-3}$ for each combination of four constellations and $R_c \in \{1/3, 1/2, 3/4, 9/10\}$. The vertical position of these 16 markers represent the resulting achievable rates and clearly show that the results follow the GMI curves. This is in good agreement with the results in [10], [35], where it was shown that the GMI can be used to predict the performance of LDPC codes for the AWGN channel. The penalties with respect to the GMI are between 5 and 15 km and are highest for high code rates and large values of M. These penalties are caused by the suboptimality of the LDPC code under consideration.

In analogy with Fig. 3, Fig. 8 shows the post-FEC BER as a function of (a) pre-FEC BER, (b) normalized MI, and (c) normalized GMI. The results for the NLSE are shown with filled markers and show that the prediction based on the GMI is excellent. Just as for TCs, the prediction based on pre-FEC BER does not always work, however, a relatively good approximation is obtained for high code rates.

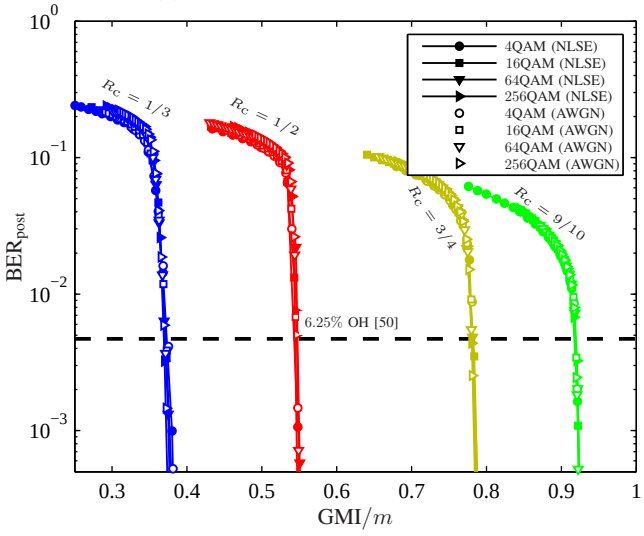
Different interleaving strategies can be used to connect the binary encoder and the mapper. These strategies include, e.g.,



(a) Post-FEC BER vs. pre-FEC BER



(b) Post-FEC BER vs. normalized MI



(c) Post-FEC BER vs. normalized GMI

Fig. 8. Post-FEC BER for LPDC codes with $R_c \in \{1/3, 1/2, 3/4, 9/10\}$ (colors) over linear and nonlinear channels and different modulation formats (markers): 4QAM, 16QAM, 64QAM, and 256QAM. The post-FEC BER is shown versus (a) pre-FEC BER, (b) normalized MI, and (c) normalized GMI. All the L-values are calculated using (29).

consecutively assigning the code bits to the mapper (as done in [36]), multiple (parallel) interleavers [66], or interleavers optimized for a particular code [9]. In this paper, an ensemble of random interleavers was considered, and thus, the *ensemble* of encoders generated by randomly permuting the code bits is considered. In general, the performance of the encoder depends on the choice of interleaver, which explains why the curves in Fig. 8 (c) are slightly more "compact" for low rates than those in [36, Fig. 5]. If a particular interleaving strategy is used instead (consecutive, parallel, optimized, etc.), the post-FEC BER prediction based on GMI is less precise, although still much better than pre-FEC BER and MI.

In Fig. 8, we also show results obtained for the AWGN channel (white markers). These results were obtained for the same modulation and coding pairs as used in the NLSE simulations and show that indeed the GMI is a robust metric to predict post-FEC BER across different channels. In particular, Fig. 8 (c) shows that the post-FEC BER predictions give the same results for both the AWGN channel and the simulations based on the NLSE. This also suggests that using a Gaussian model for the noise is quite reasonable.

All the results in Fig. 8 (c) for the NLSE were obtained for the optimal launch power. To show that the GMI prediction is also not dependent on the launch power, we study a fixed distance and vary the launch power, bringing the system deep into the nonlinear regime. As the modulation format, we choose 64QAM and based on the results in Fig. 7, we use $L = 210$ km and $R_c = 3/4$. The launch power was varied from 2.6 dBm to 12.6 dBm, giving the pre-FEC and post-FEC BER shown in Fig. 9 (a). The same post-FEC BER values are shown in Fig. 9 (b) as a function of the normalized GMI. This figure shows once again that the GMI can be used to accurately predict the post-FEC BER of SD-FEC decoders, even when the channel is highly nonlinear.

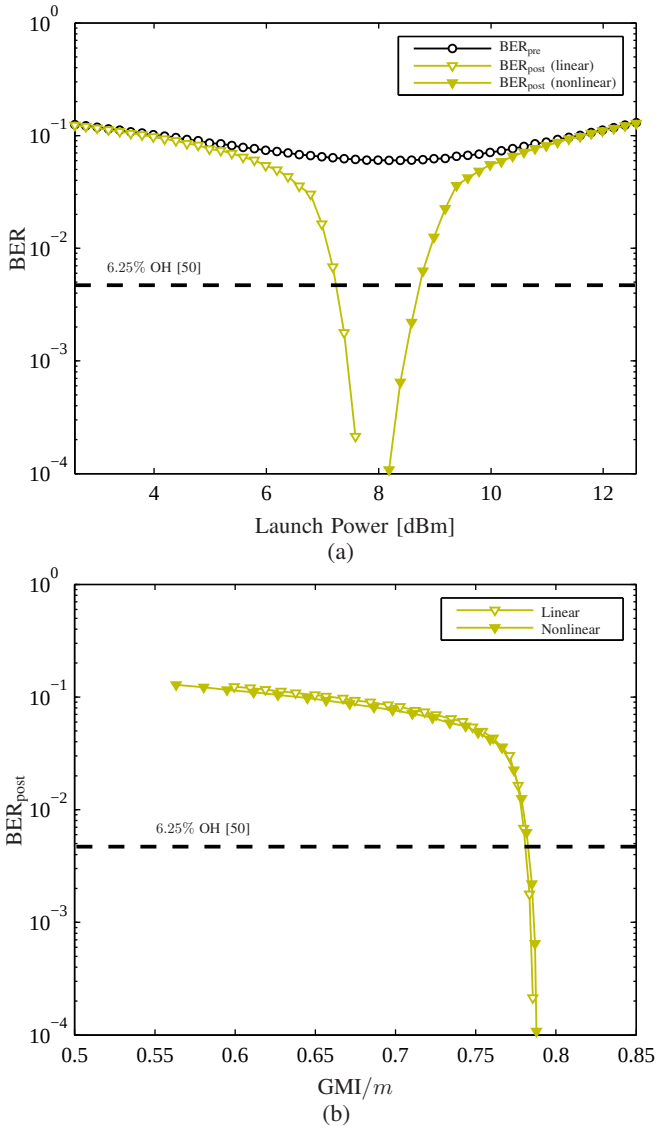


Fig. 9. (a) Pre-FEC BER and post-FEC BER as a function of the launch power for an LDPC code with $R_c = 3/4$, 64QAM, and $L = 210$ km. (b) Pre-FEC BER vs. normalized GMI in the linear (white markers) and nonlinear (filled markers) regimes.

C. Optical Channel—Experiments

To experimentally verify that the normalized GMI is an accurate predictor for post-FEC BER, the LDPC code described in Sec. II-D was implemented in a dual-polarization 64QAM Nyquist-spaced WDM transmission system. The corresponding experimental setup is illustrated in Fig. 10. A 100 kHz linewidth external cavity laser (ECL) was passed through an optical comb generator (OCG) to obtain seven frequency-locked comb lines with a channel spacing of 10.01 GHz. The eight-level drive signals required for 64QAM were generated offline in Matlab and were digitally filtered using an RRC filter with a roll-off factor of 0.1%. The resulting in-phase (I) and quadrature (Q) signals were loaded onto a pair of field-programmable gate arrays (FPGAs) and output using two digital-to-analog converters (DACs) operating at 20 Gsamples/s (2 samples/symbol). The odd and even sub-carriers were independently modulated using two complex IQ modulators, which were subsequently decorrelated before being combined and polarization multiplexed to form a Nyquist spaced 64QAM super-carrier. The recirculating loop configuration consisted of two acousto-optic switches (AOS), two EDFAs with a noise figure of 4.5 dB, an optical band-pass filter (BPF) for amplified spontaneous emission noise removal, a loop-synchronous polarization scrambler (PS) and a single 81.8 km span of Corning® SMF-28® ULL fiber.

The polarization-diverse coherent receiver had an electrical bandwidth of 70 GHz and used a second 100 kHz linewidth ECL as a local oscillator (LO). The frequency of the LO was set to coincide with the central sub-carrier of the 64QAM super-carrier and the received signals were captured using a 160 Gsamples/s real-time sampling oscilloscope with 63 GHz analog electrical bandwidth. DSP and SD-FEC decoding were subsequently performed offline in Matlab and was identical to that described in [67].

The transmission performance of the central WDM carrier was analyzed over a number of transmission distances from 81.8 km ($N_s = 1$) to 1308.8 km ($N_s = 16$) and for a number of launch powers, ranging from -18 dBm to $+2$ dBm. This resulted in a normalized GMI ranging from

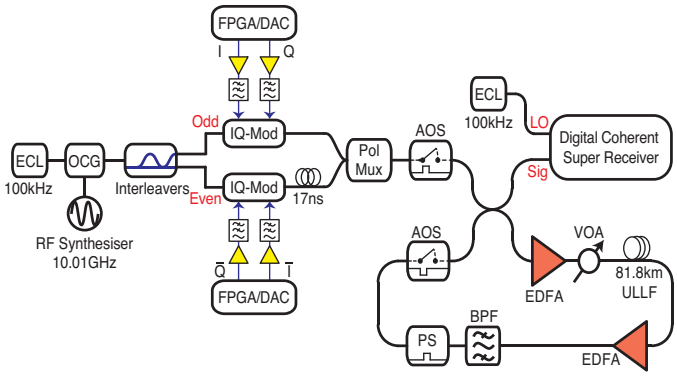


Fig. 10. Dual-polarization 64QAM Nyquist-spaced WDM transmission testbed.

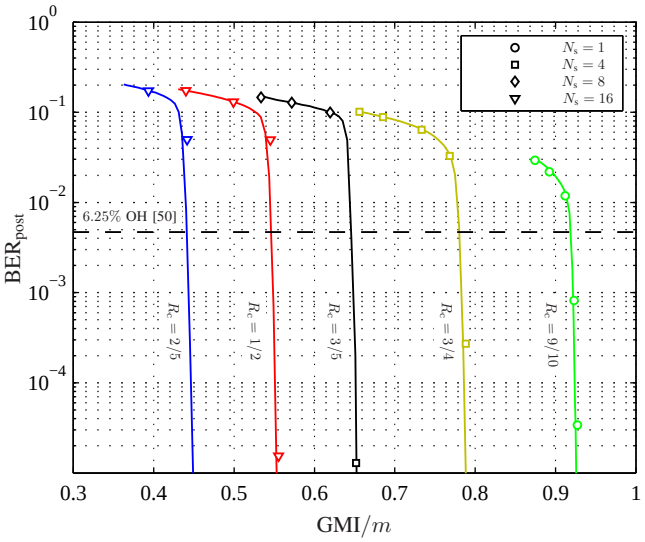


Fig. 11. Post-FEC BER for the 64QAM Nyquist spaced WDM transmission testbed with LPDC codes and $R_c \in \{2/5, 1/2, 3/5, 3/4, 9/10\}$ (colors) as a function of the normalized GMI. Experimental results for different number of spans N_s are shown with markers and AWGN results with solid lines.

0.39 to 0.93, which required adaptation of the OH in order to achieve a post-FEC BER that was below the target BER after SD-FEC decoding BER$_{post} = 4.7 \cdot 10^{-3}$. Fig. 11 illustrates the experimentally measured normalized GMI (markers) as a function of post-FEC BER, for five code rates $R_c \in \{2/5, 1/2, 3/5, 3/4, 9/10\}$. The transmission distances were $81.8, 327.2, 654.4$ and 1308.8 km, i.e., $N_s = 1, 4, 8$, and 16 spans. The simulated results obtained for an AWGN channel (i.e., the ones in Fig. 8 (c)) are displayed using solid lines. Excellent agreement between the simulated curves and the experimental points is demonstrated for all SD-FEC code rates, launch powers, and distances, even though the simulations and experiments concern entirely different channels.[11]

Each result shown with a marker in Fig. 11 corresponds to a given launch power (per channel), code rate R_c, and number of spans N_s. These results are summarized in Table II, where the launch powers are also shown. The results in Fig. 11 and Table II show that regardless of the transmit power, the

[11]The parameters of the experimental setup in this section are also different to those in Sec. IV-B.

TABLE II
SUMMARY OF RESULTS FOR THE EXPERIMENTAL SETUP IN FIG. 10. EACH ROW CORRESPONDS TO A MARKER IN FIG. 11.

Launch Power	GMI/m	BER$_{post}$	Spans N_s	Rate R_c
-18.27 dBm	0.39	$1.7 \cdot 10^{-1}$	16	2/5
-17.00 dBm	0.44	$5.0 \cdot 10^{-2}$	16	
-17.00 dBm	0.44	$1.7 \cdot 10^{-1}$	16	1/2
-15.90 dBm	0.50	$1.3 \cdot 10^{-1}$	16	
-14.80 dBm	0.55	$4.9 \cdot 10^{-2}$	16	
-13.69 dBm	0.56	$1.5 \cdot 10^{-5}$	16	
-18.12 dBm	0.53	$1.5 \cdot 10^{-2}$	8	3/5
-17.14 dBm	0.57	$1.3 \cdot 10^{-2}$	8	
-15.94 dBm	0.61	$9.9 \cdot 10^{-3}$	8	
-18.21 dBm	0.64	$1.3 \cdot 10^{-5}$	4	
-18.20 dBm	0.66	$1.0 \cdot 10^{-1}$	4	3/4
-17.20 dBm	0.69	$8.8 \cdot 10^{-2}$	4	
-16.01 dBm	0.73	$6.4 \cdot 10^{-2}$	4	
-15.0 dBm	0.77	$3.3 \cdot 10^{-2}$	4	
-0.60 dBm	0.78	$2.7 \cdot 10^{-4}$	4	
-9.17 dBm	0.87	$2.9 \cdot 10^{-2}$	1	9/10
-10.17 dBm	0.89	$2.2 \cdot 10^{-2}$	1	
-11.24 dBm	0.91	$1.2 \cdot 10^{-2}$	1	
-12.24 dBm	0.92	$8.2 \cdot 10^{-4}$	1	
-13.30 dBm	0.93	$3.4 \cdot 10^{-5}$	1	

normalized GMI can indeed be used to predict the post-FEC BER. These results can be seen as experimental validation of those presented in Fig. 9.

V. CONCLUSIONS

This paper studied the GMI as a powerful tool to predict the post-FEC BER of soft-decision FEC. The GMI was measured in experiments and simulations, and for all considered scenarios proved to be robust. The GMI can be used to predict the post-FEC BER without actually encoding and decoding data. This prediction can be done, e.g., by using the results in Table III, which shows the required normalized GMI for the codes considered in this paper. Using this table, it is correct to claim that if the measured normalized GMI in a given simulation or experiment was e.g., 0.83, there exists a concatenation of an LDPC code and a staircase code with overall rate 0.75 which gives a BER of 10^{-15} after soft-decision decoding.

The pre-FEC BER and MI were also shown to be weak predictors of the performance of soft-decision FEC for bit-wise decoders. The so-called FEC limit is, hence, an unreliable design criterion for optical communication systems with soft-decision FEC. On the other hand, the GMI was found to give very good results for all code rates, all considered modulation formats, LDPC and turbo codes, exact and approximated L-values, and for both linear and nonlinear optical transmission. We suggest to replace the "SD-FEC limit" (used for many years with hard-decision decoding and now becoming increasingly popular with soft decision) with a "GMI limit", which is relevant for modern optical communication systems.

This paper considered only noniterative binary decoding. Different results are expected if a (capacity-approaching soft-decision) nonbinary decoder or a binary decoder with iterative detection (i.e., with soft information being exchanged iteratively between the decoder and demapper) are used. In these

TABLE III
NORMALIZED GMI AND OVERALL RATE REQUIRED FOR LDPC AND TURBO CODES WITH DIFFERENT CODE RATES R_c CONCATENATED WITH A STAIRCASE CODE TO ACHIEVE A BER OF 10^{-15} AFTER DECODING.

R_c	Overall Rate	GMI LDPC	GMI Turbo
1/4	0.24	0.30	–
1/3	0.31	0.37	0.38
2/5	0.38	0.44	0.45
1/2	0.47	0.54	0.55
3/5	0.56	0.64	0.65
2/3	0.63	0.71	0.71
3/4	0.71	0.78	0.79
4/5	0.75	0.83	–
5/6	0.78	0.86	0.86
8/9	0.84	0.91	–
9/10	0.85	0.92	–

cases, we conjecture the MI to be the correct metric to predict the post-FEC BER. This comparison is left for future work.

ACKNOWLEDGMENTS

The authors would like to thank Mikhail Ivanov and Christian Häger (Chalmers University of Technology) for fruitful discussions regarding the relationship between the GMI and L-values, and Prof. Leszek Szczecinski (INRS-EMT) and Dr. Laurent Schmalen (Alcatel-Lucent Bell Labs) for fruitful discussions regarding post-FEC BER prediction. The authors would also like to thank Dr. Sergejs Makovejs and Corning Inc. for supplying the fiber used in the experimental setup.

REFERENCES

[1] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.

[2] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design*. John Wiley & Sons, 2015.

[3] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.

[4] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[5] I. B. Djordjevic, S. Sankaranarayanan, S. K. Chilappagari, and B. Vasic, "Low-density parity-check codes for 40-Gb/s optical transmission systems," *IEEE J. Quantum Electron.*, vol. 12, no. 4, pp. 555–562, July/Aug. 2006.

[6] H. Bülow and T. Rankl, "Soft coded modulation for sensitivity enhancement of coherent 100-Gbit/s transmission systems," in *Proc. Optical Fiber Communication Conference (OFC)*, San Diego, CA, Mar. 2009.

[7] H. Bülow, Ü. Abay, A. Schenk, and J. B. Huber, "Coded modulation of polarization- and space-multiplexed signals," in *Asia Communications and Photonics Conference and Exhibition (ACP)*, Shanghai, China, Nov. 2011.

[8] D. S. Millar, T. Koike-Akino, R. Maher, D. Lavery, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel, "Experimental demonstration of 24-dimensional extended Golay coded modulation with LDPC," in *Proc. Optical Fiber Communication Conference (OFC)*, San Francisco, CA, Mar. 2014.

[9] C. Häger, A. Graell i Amat, F. Brännström, A. Alvarado, and E. Agrell, "Improving soft FEC performance for higher-order modulations via optimized bit channel mappings," *Opt. Express*, vol. 22, no. 12, pp. 14 544–14 558, June 2014.

[10] A. Alvarado and E. Agrell, "Achievable rates for four-dimensional coded modulation with a bit-wise receiver," in *Proc. Optical Fiber Communication Conference (OFC)*, San Francisco, CA, Mar. 2014.

[11] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, Nov. 1997.

[12] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping for QPSK modulation," *IEE Electronics Letters*, vol. 34, no. 15, pp. 1459–1460, July 1998.

[13] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, "Soft-input soft-output modules for the construction and distributed iterative decoding of code networks," *Eur. Trans. on Telecommun.*, vol. 9, no. 2, pp. 155–172, Mar.–Apr. 1998.

[14] I. B. Djordjevic, M. Cvijetic, L. Xu, and T. Wang, "Using LDPC-coded modulation and coherent detection for ultra highspeed optical transmission," *J. Lightw. Technol.*, vol. 25, no. 11, pp. 3619–3625, Nov. 2007.

[15] H. B. Batshon, I. B. Djordjevic, L. Xu, and T. Wang, "Multidimensional LDPC-Coded modulation for beyond 400 Gb/s per wavelength transmission," *IEEE Photon. Technol. Lett.*, vol. 21, no. 16, pp. 1139–1141, Aug. 2009.

[16] H. Buelow, X. Lu, L. Schmalen, A. Klekamp, and F. Buchali, "Experimental performance of 4D optimized constellation alternatives for PM-8QAM and PM-16QAM," in *Proc. Optical Fiber Communication Conference (OFC)*, San Francisco, CA, Mar. 2014.

[17] H. Bülow and E. Masalkina, "Coded modulation in optical communications," in *Proc. Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, Mar. 2011.

[18] L. Schmalen, "Energy efficient FEC for optical transmission systems," in *Proc. Optical Fiber Communication Conference (OFC)*, San Francisco, CA, Mar. 2014.

[19] A. Puc, F. Kerfoot, A. Simons, and D. L. Wilson, "Concatenated FEC experiment over 5000 km long straight line WDM test bed," in *Proc. Optical Fiber Communication Conference (OFC)*, San Diego, CA, Feb. 1999.

[20] O. Ait Sab and V. Lemaire, "Block turbo code performances for long-haul DWDM optical transmission systems," in *Proc. Optical Fiber Communication Conference (OFC)*, Baltimore, MD, Mar. 2000.

[21] T. Mizuochi, Y. Miyata, T. Kobayashi, K. Ouchi, K. Kuno, K. Kubo, K. Shimizu, H. Tagami, H. Yoshida, H. Fujita, M. Akita, and K. Motoshima, "Forward error correction based on block turbo code with 3-bit soft decision for 10-Gb/s optical communication systems," *IEEE J. Quantum Electron.*, vol. 10, no. 2, pp. 376–386, Mar./Apr. 2004.

[22] B. Vasic and I. B. Djordjevic, "Low-density parity check codes for long-haul optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 14, no. 8, pp. 1208–1210, Aug. 2002.

[23] B. Vasic, I. B. Djordjevic, and R. K. Kostuk, "Low-density parity check codes and iterative decoding for long-haul optical communication systems," *J. Lightw. Technol.*, vol. 21, no. 2, pp. 438–446, Feb. 2003.

[24] I. B. Djordjevic and B. Vasic, "Projective geometry LDPC codes for ultralong-haul WDM high-speed transmission," *IEEE Photon. Technol. Lett.*, vol. 15, no. 5, pp. 784–786, May 2003.

[25] ITU, "Forward error correction for high bit-rate DWDM submarine systems," ITU-T Recommendation G.975.1, Tech. Rep., Feb. 2004.

[26] I. B. Djordjevic, M. Arabaci, and L. L. Minkov, "Next generation FEC for high-capacity communication in optical transport networks," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3518–3530, Aug. 2009, (Invited Paper).

[27] F. Chang, K. Onohara, and T. Mizuochi, "Forward error correction for 100 G transport networks," *IEEE Commun. Mag.*, vol. 10, no. 3, pp. S48–S55, Mar. 2010.

[28] S. Beppu, K. Kasai, M. Yoshida, and M. Nakazawa, "2048 QAM (66 Gbit/s) single-carrier coherent optical transmission over 150 km with a potential SE of 15.3 bit/s/Hz," *Opt. Express*, vol. 23, no. 4, pp. 4960–4969, Feb. 2015.

[29] D. Qian, E. Ip, M.-F. Huang, M.-J. Li, and T. Wang, "698.5-Gb/s PDM-2048QAM transmission over 3km multicore fiber," in *European Conference on Optical Communications (ECOC)*, London, UK, Mar. 2013.

[30] K. Brueninghaus, D. Astély, T. Sälzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *IEEE International Symposium on Personal, Indoor and Mobile Communications (PIMRC)*, Berlin, Germany, Sep. 2006.

[31] L. Wan, S. Tsai, and M. Almgren, "A fading-insensitive performance metric for a unified link quality model," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Las Vegas, NV, Apr. 2006.

[32] M. Franceschini, G. Ferrari, and R. Raheli, "Does the performance of LDPC codes depend on the channel?" *IEEE Trans. Commun.*, vol. 54, no. 12, pp. 2129–2132, Dec. 2006.

[33] A. Leven, F. Vacondio, L. Schmalen, S. ten Brink, and W. Idler, "Estimation of soft FEC performance in optical transmission experiments," *IEEE Photon. Technol. Lett.*, vol. 23, no. 20, pp. 1547–1549, Oct. 2011.

[34] M. Salsi, R. Rios-Muller, J. Renaudier, P. Tran, L. Schmalen, A. Ghazisaeidi, H. Mardoyan, P. Brindel, G. Charlet, and S. Bigo, "38.75 Tb/s transmission experiment over transoceanic distance," in *Proc. European Conference on Optical Communication (ECOC)*, London, UK, Sep. 2013.

[35] A. Alvarado and E. Agrell, "Four-dimensional coded modulation with bit-wise decoders for future optical communications," *J. Lightw. Technol.*, vol. 33, no. 10, pp. 1993–2003, May 2015.
[36] A. Alvarado, E. Agrell, D. Lavery, and P. Bayvel, "LDPC codes for optical channels: Is the "FEC Limit" a good predictor of Post-FEC BER?" in *Proc. Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, Mar. 2015.
[37] W. E. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge University Press, 2009.
[38] C. Jones, A. Matache, T. Tian, J. Villasenor, and R. Wesel, "The universality of LDPC codes on wireless channels," in *Military Communications Conference*, Monterey, CA, 2003, pp. 440–445.
[39] I. Sason, "On universal properties of capacity-approaching LDPC code ensembles," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 2956–2990, July 2009.
[40] I. Sason and B. Shuval, "On universal LDPC code ensembles over memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5182–5202, Aug. 2011.
[41] S. Kudekar, T. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.
[42] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2675–2692, 2007.
[43] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.
[44] M. Ivanov, F. Brännström, A. Alvarado, and E. Agrell, "On the exact BER of bit-wise demodulators for one-dimensional constellations," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1450–1459, Apr. 2013.
[45] O. Açikel and W. Ryan, "Punctured turbo-codes for BPSK/QPSK channels," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1325–1323, Sep. 1999.
[46] M. A. Kousa and A. H. Mugaibel, "Puncturing effects on turbo codes," *Proc. IEE*, vol. 149, no. 3, pp. 132–138, June 2002.
[47] J. Vogt and A. Finger, "Improving the max-log-MAP turbo decoder," *IEEE Electronic Letters*, vol. 36, no. 23, pp. 1937–1939, Nov. 2000.
[48] ETSI, "Digital video broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications (DVB-S2)," ETSI, Tech. Rep. ETSI EN 302 307 V1.2.1 (2009-08), Aug. 2009.
[49] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Error probability analysis of bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 262–271, Jan. 2006.
[50] L. M. Zhang and F. R. Kschischang, "Staircase codes with 6% to 33% overhead," *J. Lightw. Technol.*, vol. 32, no. 10, pp. 1999–2001, May 2014.
[51] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
[52] 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.
[53] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. M. J. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.
[54] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6650–6672, Oct. 2011.
[55] A. Alvarado, F. Brännström, E. Agrell, and T. Koch, "High-SNR asymptotics of mutual information for discrete constellations with applications to BICM," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1061–1076, Feb. 2014.
[56] A. Alvarado, F. Brännström, and E. Agrell, "High SNR bounds for the BICM capacity," in *IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011.
[57] L. Peng, "Fundamentals of bit-interleaved coded modulation and reliable source transmission," Ph.D. dissertation, University of Cambridge, Cambridge, UK, Dec. 2012.
[58] G. Böcherer, "Probabilistic signal shaping for bit-metric decoding," in *IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI, July 2014.
[59] J. Jaldén, P. Fertl, G., and Matz, "On the generalized mutual information of BICM systems with approximate demodulation," in *IEEE Information Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010.
[60] T. Nguyen and L. Lampe, "Bit-interleaved coded modulation with mismatched decoding metrics," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 437–447, Feb. 2011.
[61] L. Szczecinski, "Correction of mismatched L-values in BICM receivers," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3198–3208, Nov. 2012.
[62] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2-QAM: Closed-form approximations and applications," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2071–2079, July 2009.
[63] B. P. Smith and F. R. Kschischang, "Future prospects for FEC in fiber-optic communications," *IEEE J. Quantum Electron.*, vol. 16, no. 5, pp. 1245–1257, Sep./Oct. 2010.
[64] M. Ivanov, C. Häger, F. Brännström, A. Graell i Amat, A. Alvarado, and E. Agrell, "On the information loss of the max-log approximation in BICM systems," Aug. 2014, available at http://arxiv.org/abs/1408.2214.
[65] C. R. Menyuk, "Nonlinear pulse propagation in birefringent optical fibers," *IEEE J. Quantum Electron.*, vol. 23, no. 2, pp. 174–176, Feb. 1987.
[66] A. Alvarado, E. Agrell, L. Szczecinski, and A. Svensson, "Exploiting UEP in QAM-based BICM: Interleaver and code design," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 500–510, Feb. 2010.
[67] R. Maher, T. Xu, L. Galdino, M. Sato, A. Alvarado, K. Shi, S. J. Savory, B. C. Thomsen, R. I. Killey, and P. Bayvel, "Spectrally shaped DP-16QAM super-channel transmission with multi-channel digital back-propagation," *Sci. Rep.*, vol. 5, pp. 1–8, Feb. 2015, article number: 8214.

Alex Alvarado (S'06–M'11–SM'15) was born in Quellón, on the island of Chiloé, Chile. He received his Electronics Engineer degree (Ingeniero Civil Electrónico) and his M.Sc. degree (Magíster en Ciencias de la Ingeniería Electrónica) from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2003 and 2005, respectively. He obtained the degree of Licentiate of Engineering (Teknologie Licentiatexamen) in 2008 and his PhD degree in 2011, both of them from Chalmers University of Technology, Gothenburg, Sweden.

Dr. Alvarado is currently a Senior Research Associate at the Optical Networks Group, University College London, United Kingdom. In 2012–2014 he was a Marie Curie Intra-European Fellow at the University of Cambridge, United Kingdom, and during 2011–2012 he was a Newton International Fellow at the same institution. In 2008, he was holder of the Merit Scholarship Program for Foreign Students, granted by the Ministère de l'Éducation, du Loisir et du Sports du Québec. Dr. Alvarado is a recipient of the 2009 IEEE Information Theory Workshop Best Poster Award, the 2013 IEEE Communication Theory Workshop Best Poster Award, and the 2015 IEEE Transaction on Communications Exemplary Reviewer Award. His general research interests are in the areas of digital communications, coding, and information theory.

Erik Agrell (M'99–SM'02) 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 signals and systems research area. He is a Visiting Professor at University College London since 2014. 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 is an Associate Editor for the IEEE Transactions on Communications since 2012. 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, and the 2013 Chalmers Supervisor of the Year Award.

Domaniç Lavery (S'09–M'13) received the M.Phys. degree in theoretical physics from the University of Durham, Durham, UK, in 2009, and the Ph.D. degree in electronic and electrical engineering from University College London (UCL), London, UK, in 2013.

His doctoral research focused on the use of digital coherent transceivers and their application to spectrally efficient, high-capacity passive optical networks. He is currently continuing his work with the Optical Networks Group at UCL as a Research Associate, investigating techniques for maximizing channel capacity in nonlinear fiber transmission.

Dr. Lavery is a member of the Optical Access Systems and Wireless Backhaul Networks subcommittee for the Optical Fiber Communication Conference (OFC, 2015-2016) and was a member of the technical programme committee for the 20th European Conference on Networks and Optical Communications (NOC). He was awarded the IEEE Photonics Society Graduate Student Fellowship in 2012 and the Marconi Societys Paul Baran Young Scholar Award in 2013.

Robert Maher (M'09–SM'15) received the B.Eng. and Ph.D. degrees in electronic engineering from Dublin City University, Ireland, in 2005 and 2009, respectively. His doctoral research focused on the development and characterization of cost-efficient wavelength tunable transmitters suitable for reconfigurable agile optical networks.

In 2010, he received an IRC INSPIRE Marie Curie Fellowship and joined the Optical Networks Group at University College London (UCL), London, U.K. In 2013, he was appointed as a Senior Research Associate on the UNLOC EPSRC program grant, within the Optical Networks Group at UCL. His current research interests include spectrally efficient long-haul transmission for coherent optical networks, fiber nonlinearity mitigation techniques, and dynamic optical networking.

Dr. Maher is a Member of the Marie Curie Fellows Association.

Polina Bayvel (S'87–M'89–SM'00–F'10) received her B.Sc. (Eng) and Ph.D. degrees in Electronic & Electrical Engineering from University of London, UK, in 1986 and 1990, respectively. In 1990, she was with the Fiber Optics Laboratory, General Physics Institute, Moscow (Russian Academy of Sciences), under the Royal Society Postdoctoral Exchange Fellowship. She was a Principal Systems Engineer with STC Submarine Systems, Ltd., London, UK, and Nortel Networks (Harlow, UK, and Ottawa, ON, Canada), where she was involved in the design and planning of optical fibre transmission networks. During 1994-2004, she held a Royal Society University Research Fellowship at University College London (UCL), and in 2002, she became a Chair in Optical Communications and Networks. She is currently the Head of the Optical Networks Group, UCL which she also set up in 1994. She has authored or co-authored more than 300 refereed journal and conference papers. Her research interests include wavelength-routed optical networks, high-speed optical transmission, and the study and mitigation of fibre nonlinearities.

Prof. Bayvel is a Fellow of the Royal Academy of Engineering (FREng.), the Optical Society of America, the UK Institute of Physics, and the Institute of Engineering and Technology. She was the recipient of the Royal Society Wolfson Research Merit Award (2007-2012), 2013 IEEE Photonics Society Engineering Achievement Award and 2014 Royal Society Clifford Patterson Prize Lecture and Medal.