Performance Prediction of Nonbinary Forward Error Correction in Optical Transmission Experiments

Laurent Schmalen, Alex Alvarado, and Rafael Rios-Müller

(Top Scored)

***Abstract*—In this paper, we compare different metrics to predict the error rate of optical systems based on nonbinary forward error correction (FEC). It is shown that an accurate metric to predict the performance of coded modulation based on nonbinary FEC is the mutual information. The accuracy of the prediction is verified in a detailed example with multiple constellation formats and FEC overheads, in both simulations and optical transmission experiments over a recirculating loop. It is shown that the employed FEC codes must be universal if performance prediction based on thresholds is used. A tutorial introduction into the computation of the thresholds from optical transmission measurements is also given.**

***Index Terms*—Bit error rate, coded modulation, forward error correction, generalized mutual information, mutual information, performance prediction, symbol error rate.**

I. Introduction and Motivation

MANY optical transmission experiments do not include forward error correction (FEC). The reasons for this are that often, FEC development is still ongoing, or FEC developers are physically remote from the experiment. Often, researchers would also like to reuse experimental data obtained in expensive optical transmission experiments to evaluate the performance of different FEC schemes, without needing to redo the transmission experiment and/or signal processing. Therefore, *thresholds* are commonly used to decide whether the bit error rate (BER) after FEC decoding is below the required target BER, which can be in the range of 10^{-13} to 10^{-15}. The most commonly used threshold in the optical communications literature is the pre-FEC BER.

The use of thresholds is also very convenient in practice because very low post-FEC BER values are hard to estimate. The conventional design strategy has therefore been to experimentally demonstrate (or simulate) systems without FEC encoding and decoding, and to optimize the design for a much higher BER value, the so-called "FEC limit" or "FEC threshold." This approach relies on the strong assumption that a certain BER without coding can be reduced to the desired post-FEC BER by previously verified FEC implementations, regardless of the system under consideration.

Using pre-FEC BER thresholds is very popular in the literature and has been used for example in the record experiments based on 2048 quadrature amplitude modulation (QAM) for single-core [2] and multi-core [3] fibers. This threshold indeed gives accurate post-FEC BER predictions if three conditions are satisfied. First, bit-level interleaving must be used to guarantee independent bit errors. Second, the FEC under consideration must be binary and universal, and lastly, the decoder is based on hard decisions (bits) rather than soft decisions. Recently, however, it was shown in [4], [5] that the pre-FEC BER fails at predicting the post-FEC BER of binary *soft-decision* FEC. This was shown for both turbo codes and low-density parity-check (LDPC) codes, in the linear and nonlinear regimes, and in both simulations and optical experiments. Furthermore, [4] also showed that a better predictor in this case is the generalized mutual information (GMI)[1] [6, Sec. 3], [7, Sec. 4.3], [8], [9] and suggested to replace the pre-FEC BER threshold by a "GMI threshold."

The rationale for using the GMI as a metric to characterize the performance of binary soft-decision FEC is that the GMI is an achievable information rate (AIR) for bit-interleaved coded modulation (BICM) [6], [7], often employed as a pragmatic approach to coded modulation (CM). For square QAM constellations, BICM operates close to capacity with moderate effort, and thus, it is an attractive CM alternative. However, for most nonsquare QAM constellations, BICM results in unavoidable performance penalties. For these modulation formats, other coded modulation (CM) schemes such as nonbinary (NB) FEC [10] and multi-level coding with multi-stage decoding [11] can be advantageous. Furthermore, BICM is not expected to be the most complexity-efficient coded modulation scheme for short reach and metro optical communications with higher order modulation. The reason is that the digital signal processing (DSP) implementation needs to work at the transmission baud rate, but the FEC decoder needs to operate at m times the DSP

Manuscript received June 1, 2016; revised August 8, 2016; accepted September 9, 2016. Date of publication September 15, 2016; date of current version February 22, 2017. The work of the author L. Schmalen was supported by the CELTIC EUREKA through SENDATE-TANDEM Project C2015/3-2, which is partly funded by the German BMBF Project 16KIS0450K. The work of the author A. Alvarado was supported by the Engineering and Physical Sciences Research Council, U.K. through project UNLOC under Grant EP/J017582/1. Parts of this paper have been presented at the 2016 Optical Fiber Communication Conference (OFC), Anaheim, CA, USA, Mar. 2016 in paper M2A.2 [1].

L. Schmalen is with Nokia Bell Labs, Stuttgart 70435, Germany (e-mail: laurent.schmalen@nokia.com).

A. Alvarado is with the the Optical Networks Group, Department of Electronic & Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: alex.alvarado@ieee.org).

R. Rios-Müller is with Nokia Bell Labs, Paris-Saclay, Nozay 91620, France (e-mail: rafael.rios_muller@alcatel-lucent.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2016.2609932

[1] Also known as the BICM capacity or parallel decoding capacity.

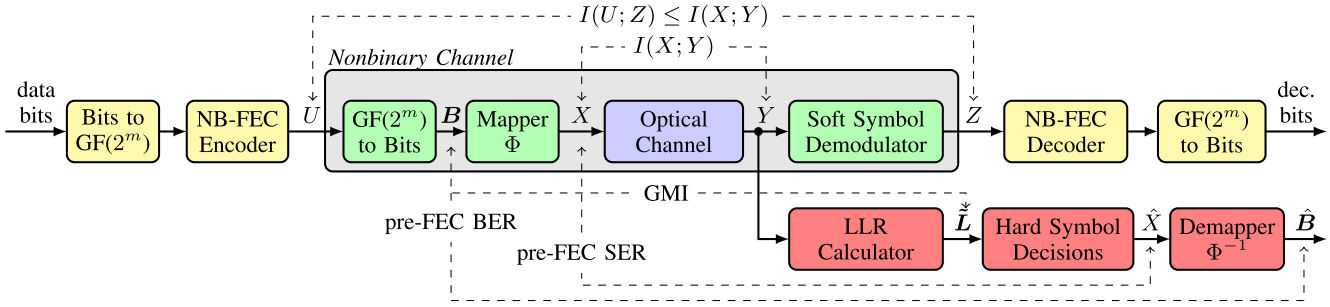

Fig. 1. System model of optical transmission based on NB-CM and the measurement of various system parameters.

rate, if 2^m-ary higher order modulation formats are used. For these applications, multi-level coding [12], [13] or NB-FEC may be good candidates and for these, the throughput is in the same order as for the DSP. Although most nonbinary FEC schemes are considerably more complex to implement than their binary counterparts, recent advances [14], [15] show that very low-complexity nonbinary FEC schemes for higher order constellations can be implemented. A numerically stable algorithm with high parallelism was presented in [16].

In this paper, we investigate the performance prediction of NB soft-decision FEC (NB-FEC) and show that an accurate threshold in this case is the mutual information (MI) [17]. The MI was previously introduced in [18] to assess the performance of differentially encoded quaternary phase shift keying and was shown to be a better performance indicator than the pre-FEC BER. The use of MI as a post-FEC BER predictor for capacity-approaching nonbinary FEC was also conjectured in [4, Sec. V] and was previously suggested in [19], [20] in the context of wireless communications.

The main contribution of this paper is to show that the MI is an accurate threshold for a CM scheme based on NB LDPC codes. This is verified in both an additive white Gaussian noise (AWGN) simulation and in two optical experiments using 8-QAM constellations. We show that the MI allows us to accurately predict the post-FEC performance of NB LDPC schemes and also show that other commonly used thresholds (such as pre-FEC BER, pre-FEC symbol error rate (SER) and bit-wise GMI) fail in this scenario.

This paper is organized as follows. In Section II we describe the system model we use and lay down some information theory preliminaries. Afterwards, in Section III we show what thresholds we should use to predict the performance of NB FEC schemes. In Section IV, we verify our predictors with a simulation example, a back-to-back experiment and a transmission experiment over a recirculating loop. Finally, in Section V, we discuss code universality and give guidelines for using the proposed thresholds.

II. System Model and Preliminaries

A. System Model

Fig. 1 shows the NB-CM scheme under consideration. The data bits are mapped to NB symbols from GF(2^m) using a one-to-one (i.e., invertible) mapping function, then encoded by an NB-FEC with rate R, and then mapped to D-dimensional constellation symbols from the set $\mathcal{S} := \{s_1, \ldots, s_M\}$, where $|\mathcal{S}| = 2^m = M$ and $s_i \in \mathbb{R}^D$. Frequently, $D = 2$ (with complex symbols), but in optical communications, also $D = 4$ [8], [9], [21], [22] and $D = 8$ [23], [24] are used. As will become obvious later, the mapping to symbols is shown in two stages in Fig. 1, namely first mapping the NB symbols $U \in \{1, 2, \ldots, M\}$ to bit patterns $\boldsymbol{B}$ of m bits, and mapping these to constellations symbols $X \in \mathcal{S}$. In some cases, we require the combination of bit mapper and mapper Φ, which we denote by $\phi(i) = s_i$ and which maps an integer i to a modulation symbol s_i.

The constellation symbol $s_i \in \mathcal{S}$ is transmitted with *a priori* probability $P(X = s_i) := \lambda_i$ through an "optical channel."[2] Most communication systems transmit equiprobable symbols, i.e., $\lambda_i = 1/M$, $\forall i$. However, in the case of probabilistic shaping [25]–[27], the probabilities of occurrence of the symbols may differ. The optical channel[3] takes a sequence of N_m constellation symbols $\boldsymbol{x}_1^{N_m} := (x[1], x[2], \ldots, x[N_m])$ and maps them to a waveform $w(t)$ by means of a pulse shaping function $\rho(t)$ with

$$w(t) = \sum_{\kappa=1}^{N_m} x[\kappa] \cdot \rho(t - \kappa T_s)$$

with T_s being the symbol period and κ the discrete-time index. The optical channel further includes digital-to-analog converters (DACs), filtering, transmission including amplification, analog-to-digital converters (ADCs), and DSP to remove effects of chromatic dispersion, polarization mode dispersion, polarization rotation, phase noise, frequency offset, etc. It further includes matched filtering, equalization and possibly (de-)interleaving.

At the receiver, for each sampled symbol $y[\kappa]$, the soft symbol demodulator (see Fig. 1) computes M scaled likelihoods (which are proportional to the *a posteriori* probabilities) $q_{Y|X}(y|s_i)\lambda_i$, where $q_{Y|X}(y|s_i)$ is a function that depends on the received D-dimensional sampled symbol y and the constellation symbol $s_i \in \mathcal{S}$. These scaled likelihoods are passed to an NB-FEC decoder. Note that usually, for numerical reasons, a vec-

[2]We use upper-case letters (e.g., X) to denote random variables and lower-case letters (e.g., x) to denote realizations of this random variable. We use boldface upper-case letters (e.g., $\boldsymbol{X}$) to denote sequences of random variables and boldface lower-case letters (e.g., $\boldsymbol{x}$) for their realizations. Sets are denoted by calligraphic letters (e.g., $\mathcal{S}$). $\|\cdot\|$ is used to denote the L_2 norm.

[3]Also referred to as "discrete-time (noisy) channel" in [28].

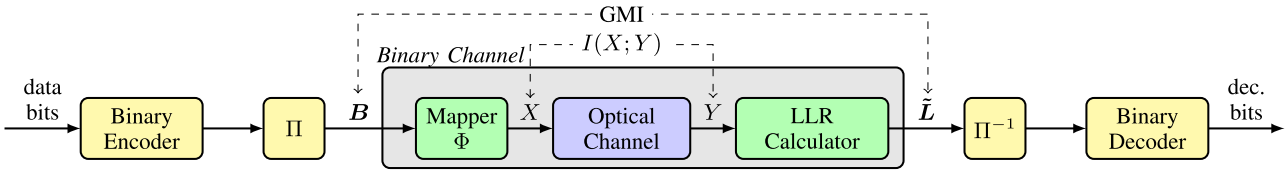

Fig. 2. System model of optical transmission based on BICM.

tor of $M-1$ nonbinary log-likelihood ratios (LLRs) is computed for each D-dimensional received symbol y instead. These (nonbinary) LLRs[4] are given by

$$L_i(y) = \ln\left(\frac{q_{Y|X}(y|s_i)}{q_{Y|X}(y|s_1)}\right) + \ln\left(\frac{\lambda_i}{\lambda_1}\right),\ \forall i \in \{2,\ldots,M\}. \tag{1}$$

Ideally, the receiver knows the (averaged) optical channel transition probability density function (PDF) $p_{Y|X}(y|s_i)$, applies sufficiently long interleaving, and sets $q_{Y|X}(y|x) = p_{Y|X}(y|x)$ in (1). Usually, however, the exact channel transition PDF is not known at the receiver, or the computation of the LLRs is too involved using the true PDF, which is why often approximations are used. In this case $q_{Y|X}(y|x) \neq p_{Y|X}(y|x)$, and thus, we say that the receiver is *mismatched* [30]. Often, for instance, the (multivariate) Gaussian PDF is assumed at the receiver, i.e., $q_{Y|X}(y|s_i) = q_{\mathsf{awgn}}(y|s_i)$, where

$$q_{\mathsf{awgn}}(y|s_i) := \frac{\exp\left(-\frac{1}{2}(y-s_i)^T\boldsymbol{\Sigma}^{-1}(y-s_i)\right)}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}}.$$

In [22], different approximations are compared for $D=4$ and it was found that the circularly symmetric Gaussian approximation with diagonal covariance matrix $\boldsymbol{\Sigma}$ reliably approximates the true PDF unless the input power is increased to very high levels. Besides, the Gaussian PDF has also been shown to be a good approximation for the true PDF in case of uncompensated fiber links with coherent reception [31]. Furthermore, using a Gaussian PDF also simplifies the numerical computation of the LLRs.

A predominant case is $D=2$ (e.g., QAM constellations detected independently in each polarization) with circularly symmetric noise (diagonal $\boldsymbol{\Sigma}$) and variance σ_n^2 per dimension. This is the case on which we focus on this paper and which is also dominant in coherent long-haul dispersion uncompensated links [22]. In this case

$$q_{\mathsf{awgn}}(y|s_i)\big|_{D=2} = \frac{1}{2\pi\sigma_n^2}\exp\left(-\frac{\|y-s_i\|^2}{2\sigma_n^2}\right). \tag{2}$$

Assuming equally likely symbols ($\lambda_i = 1/M$), the LLRs in (1) are given by

$$L_i(y) = \frac{1}{2\sigma_n^2}(\|y-s_1\|^2 - \|y-s_i\|^2). \tag{3}$$

After LLR computation, the NB soft-decision FEC decoder (e.g., a nonbinary LDPC decoder) takes these LLRs and estimates the transmitted NB symbols, which are later converted into decoded bits. Here we only assume that the nonbinary FEC is matched to the constellation, i.e., each nonbinary symbol of the FEC code can be mapped to $m = \log_2(M)$ bits. This allows us to consider nonbinary LDPC codes defined over either the Galois field GF(2^m) or the ring $\mathbb{Z}_M$ of integers modulo M. We further assume that *soft decision* decoding is carried out, see, e.g., [16]. For other, low complexity versions of that algorithm, we refer the interested reader to the references in [16].

B. Bit-Interleaved Coded Modulation (BICM)

In optical communications, often the pragmatic BICM scheme is used. The system model of BICM is shown in Fig. 2. We only describe a simplified version here. For more details, we refer the interested reader to [7], [11] and references therein. BICM is based on a binary FEC code. The binary output of the FEC encoder is interleaved[5] by a permutation function Π. The resulting interleaved bit stream $\boldsymbol{B}$ is then mapped to modulation symbols X using the mapper Φ described above.

At the receiver, we use a bit metric decoder (BMD) to compute LLRs $\tilde{\boldsymbol{L}}$ for the individual bits of the bit stream $\boldsymbol{B}$. In Fig. 2, the BMD is denoted *LLR Calculator*. The LLRs computed by the BMD are given by

$$\tilde{L}_i(y) = \log\left(\frac{\sum_{s\in\mathcal{S}_{0,i}} q_{Y|X}(y|s)\lambda_{\phi^{-1}(s)}}{\sum_{s\in\mathcal{S}_{1,i}} q_{Y|X}(y|s)\lambda_{\phi^{-1}(s)}}\right),\ \forall i \in \{1,\ldots,m\} \tag{4}$$

where $\mathcal{S}_{b,i}$ is the set of constellation symbols where the i-bit of the binary label takes on the value b. In the practically dominant case with equiprobable symbols ($\lambda_i = 1/M$), we get

$$\tilde{L}_i(y) = \log\left(\frac{\sum_{s\in\mathcal{S}_{0,i}} q_{Y|X}(y|s)}{\sum_{s\in\mathcal{S}_{1,i}} q_{Y|X}(y|s)}\right)$$

The stream of LLRs $\tilde{\boldsymbol{L}}$ is then de-interleaved by the inverse permutation Π^{-1} and then fed to a conventional soft-decision binary FEC decoder.

The comparison of (4) with (1) clearly shows the difference between nonbinary CM and BICM. In the nonbinary case, we compute a vector of LLRs containing $M-1$ values for each channel observation Y. In contrast, for BICM, we only compute $m = \log_2(M)$ LLRs per channel observation. Clearly, there is a compression of information which is available for the FEC decoder. Fascinatingly, the loss of information from this compression can be made negligible in many practical cases, e.g., with square QAM constellations and Gray mapping [7], [11].

[4]Strictly speaking, the quantities in (1) are log-*a posteriori* probability (log-APP) ratios. However, in the FEC literature, the log-APP ratios are typically also called LLRs, which is why we follow this latter convention here (see also [29, Sec. 4.5.3] and [7, p. 58]).

[5]Often, the interleaver is considered to be part of the FEC encoder, for instance if random LDPC codes are used.

The loss of information may however become important for other constellations.

C. FEC Universality

When assessing and comparing the performance of different modulation formats and different transmission scenarios (e.g., fiber types, modulators, converters, etc.) based on *thresholds*, it is important to understand the concept of FEC *universality*. A pair of FEC code and its decoder are said to be *universal* if the performance of the code (measured in terms of post-FEC BER or SER) does not depend on the nonbinary channel (with input U and output Z when referring to Fig. 1), provided that the channel has a fixed mutual information $I(U;Z)$.

When we refer to "the channel", we consider the whole transmission chain between the NB-FEC encoder output U and the decoder input Z including modulation and demodulation, DSP, ADCs and DACs, optical transmission and amplification including noise. We say that the channel changes if any of the components in the chain between U and Z changes. This can be for instance the noise spectrum, the optical signal-to-noise ratio (SNR), but also the modulation format or the DSP algorithms. We provide a rigorous definition of universality later in Section V.

Unfortunately, not much is known about the universality of practical coding schemes. It is conjectured that many practical (binary) LDPC codes are approximately universal [32] which has been shown to be asymptotically true under some relatively mild conditions [33]. Guidelines for designing LDPC codes that show good universality properties are highlighted in [34]. The class of spatially coupled LDPC codes, recently investigated for optical communications [35], has been shown to be asymptotically universal [36]. An example of *non*-universal coding schemes are the recently proposed, capacity-achieving polar codes [37], which need to be redesigned for every different channel. Most of these results are for binary codes and even less is known for nonbinary codes.

Although most practical LDPC codes are asymptotically universal, we wish to emphasize a word of caution: practical, finite-length realizations of codes may only be approximately universal. For instance, [32, Fig. 3] reveals that for some practical LDPC codes, the performance at a BER of 10^{-4} significantly differs for different channels. This difference is expected to be even larger at very low BERs due to the different slopes of the curves. We will address this difference in detail in Section V.

D. Channel Capacity and Mutual Information

Consider an information stable, discrete-time channel with memory [38]–[40], which is characterized by the sequence of PDFs $p_{\boldsymbol{Y}_1^N|\boldsymbol{X}_1^N}(\boldsymbol{y}_1^N|\boldsymbol{x}_1^N)$, for $N=1,2,\ldots$. The maximum rate at which reliable transmission over such a channel is possible is defined by the *channel capacity* [38]–[40]

$$C := \lim_{N\to\infty} \sup_{p_{\boldsymbol{X}_1^N}} \frac{1}{N} I(\boldsymbol{X}_1^N;\boldsymbol{Y}_1^N) \tag{5}$$

where the maximization is over $p_{\boldsymbol{X}_1^N}(\cdot)$, which is the PDF of the sequence $\boldsymbol{X}_1^N = (X[1], X[2], \ldots, X[N])$ under a given input constraint (e.g., power constraint). For a fixed $p_{\boldsymbol{X}_1^N}(\cdot)$, the mutual information (MI) between the input sequence $\boldsymbol{X}_1^N$ and the output sequence $\boldsymbol{Y}_1^N$ is given by

$$I(\boldsymbol{X}_1^N;\boldsymbol{Y}_1^N) = \mathbb{E}_{p_{\boldsymbol{X}_1^N,\boldsymbol{Y}_1^N}}\left\{\log_2 \frac{p_{\boldsymbol{Y}_1^N|\boldsymbol{X}_1^N}(\boldsymbol{Y}_1^N|\boldsymbol{X}_1^N)}{p_{\boldsymbol{Y}_1^N}(\boldsymbol{Y}_1^N)}\right\}$$

where $\mathbb{E}_{p_R}\{f(R)\} := \int_{\mathrm{dom}(R)} p_R(r) f(r) \mathrm{d}r$ denotes expectation with respect to a random variable R.

The capacity C in (5) is the maximum information rate that can be achieved for any transmission system, requiring carefully optimized, infinitely long input sequences. Usually, in most of today's systems, the channel input sequence is heavily constrained (e.g., by the use of QAM constellations) to simplify the transceiver design. Furthermore, often symbol sequences with independent and identically distributed (IID) elements are used such that we have

$$p_{\boldsymbol{X}_1^N}(\boldsymbol{x}_1^N) = \prod_{i=1}^{N} P_X(x[i]) = \prod_{i=1}^{N} \lambda_{\phi^{-1}(x[i])}. \tag{6}$$

IID symbol sequences are obtained if a memoryless mapper is used (as we do in this paper, see, e.g., Φ in Fig. 1) and if sufficiently long interleaving is applied after FEC encoding. Under these conditions, an achievable information rate (AIR) is given by

$$I_{\mathrm{mem}} = \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{X}_1^N;\boldsymbol{Y}_1^N) \le C \tag{7}$$

which is a lower bound to the capacity C due to the constraints imposed on the transmitted sequences. In the remainder of this paper, we limit ourselves to IID channel input sequences generated via (6).

The numerical evaluation of the MI in (7) is in general not practical. The reasons are as follows: First, numerically evaluating $I(\boldsymbol{X}_1^N;\boldsymbol{Y}_1^N)$ is hard, even for for relatively short input and channel output sequences (small memory lengths N). Second, most of today's transceivers do not exploit memory but instead use long interleavers to remove all effects of memory to keep decoding simple with symbol-by-symbol detection. Hence, it would not be fair to provide thresholds based on memory, which give a performance that could be achieved at some point in the future, provided that all memory is adequately exploited at the transceiver. Instead, we neglect all memory effects and obtain thresholds that indicate a performance *achievable* with today's systems.

Therefore in this paper, we focus on symbol-by-symbol detection (see Fig. 1). Under these constraints, we can further lower bound the MI in (7) (see [41, Sec. III-F] for an in-depth proof) by employing a memoryless channel transition PDF $p_{Y|X}(\cdot|\cdot)$ that is obtained by averaging the true channel PDF. This approach gives

$$I(X;Y) = \mathbb{E}_{p_{X,Y}}\left\{\log_2 \frac{p(Y|X)}{p(Y)}\right\} \le I_{\mathrm{mem}} \le C \tag{8}$$

or equivalently

$$I(X;Y) = \sum_{i=1}^{M} \lambda_i \int\limits_{y\in\mathbb{R}^D} p_{Y|X}(y|s_i) \log_2 \left(\frac{p_{Y|X}(y|s_i)}{\sum_{j=1}^{M} p_{Y|X}(y|s_j)\lambda_j} \right) \mathrm{d}y. \tag{9}$$

Note that $I(X;Y)$ is an AIR for systems employing optimum decoding, i.e., when the LLR computation uses $q_{Y|X}(y|x) = p_{Y|X}(y|x)$, and if sufficiently long symbol-wise interleaving is applied (within the equivalent "optical channel") and sufficiently long capacity-achieving nonbinary FEC codes are used.

III. Thresholds for Nonbinary FEC

Based on the discussion in Section II-C, here we propose to use the MI as performance thresholds for NB-FEC. After a discussion on how to compute these thresholds, we describe some other commonly used thresholds.

A. Thresholds Based on Mutual Information

In order to estimate the performance of NB-FEC, motivated by the universality argument in Section II-C, we would like to use the MI $I(U;Z)$ as performance threshold. $I(U;Z)$ is the MI between the FEC encoder output U and FEC decoder input Z (see Fig. 1) and characterizes the nonbinary channel. Unfortunately, the MI $I(U;Z)$ is not easy to compute immediately, which is why we define a threshold that is directly related to the input X and output Y of the optical transmission experiment, to which we usually have access. This also allows us to avoid including soft symbol demodulation in the transmission experiment.

In the previous section, we have seen that I_{mem} is a maximum AIR if all memory effects are taken into account and is an upper bound on $I(X;Y)$, which is an AIR under optimum decoding with an averaged channel PDF. As a consequence of the data processing inequality, we have

$$I_{\text{mem}} \geq I(X;Y) \overset{(a)}{\geq} I(U;Z)$$

where we have equality in (a) only in some special cases described below. Due to this inequality, we cannot always directly use $I(X;Y)$ as a proxy for estimating $I(U;Z)$. We resort to the theory of mismatched decoding [42][30] and propose to use $\underline{I}(X;Y)$ as estimate of $I(U;Z)$, where

$$\underline{I}(X;Y) := \sup_{\nu\geq 0} \mathbb{E}_{p_{X,Y}} \left\{ \log_2 \left(\frac{[q_{Y|X}(Y|X)]^\nu}{\sum_{j=1}^{M} \lambda_j [q_{Y|X}(Y|s_j)]^\nu} \right) \right\}. \tag{10}$$

We have $I(X;Y) \geq \underline{I}(X;Y) \geq I(U;Z)$, where the second inequality is due to [30], [42]. However, we found in numerical simulations and in transmission experiments that, in the context of optical communications, $\underline{I}(X;Y) \approx I(U;Z)$. Hence, we use $\underline{I}(X;Y)$ as an accurate estimate of $I(U;Z)$ and of the NB-FEC performance.

Even (10) is demanding to evaluate in general, as the expectation is taken over $P_{Y,X}(y,x) = p_{Y|X}(y|x)\lambda_{\phi^{-1}(x)}$, which is often not known. However, we can replace the expectation in (10) by the empirical average, as done for instance in [25, Sec. III]. We denote this empirical approximation of $\underline{I}(X;Y)$ by I_{NB}, which can be computed from an optical transmission experiment with a measurement database of N_m measured values $x[\kappa] \in \mathcal{S}$ and their corresponding received $y[\kappa]$ by

$$I_{\text{NB}} := \frac{1}{N_m} \sup_{\nu\geq 0} \sum_{\kappa=1}^{N_m} \log_2 \left(\frac{[q_{Y|X}(y[\kappa]|x[\kappa])]^\nu}{\sum_{j=1}^{M} \lambda_j [q_{Y|X}(y[\kappa]|s_j)]^\nu} \right) \tag{11}$$

where $q_{Y|X}(y|x)$ is the same PDF used for computing the LLRs in (1), e.g., the $D = 2$-dimensional Gaussian PDF. The variance of this distribution can for instance be estimated from the measurement database (or a subset thereof), see, e.g., [25, Sec. III]. Later, in Example 2, we show how we can jointly estimate the MI and the noise variance, avoiding an extra variance estimator. As the optimization in (10) and (11) is over a strictly unimodal ($\cap$-convex) function in ν [7, Thm. 4.22], the maximization can be efficiently carried out using, e.g., the Golden section search [43].

B. Detailed Description of the Proposed Threshold $\underline{I}(X;Y)$

In the following, we describe in detail the steps that lead us to the performance metric in (11) starting from $I(U;Z)$. The remainder of this section may be skipped in a first reading. The input Z to the FEC decoder consists of vectors of $M-1$ dimensional LLRs, whose distributions are hard to estimate, especially if M becomes large. Therefore, we would like to relate $I(U;Z)$ to X and Y, to which we have immediately access as input and output parameters of the optical transmission experiment. Using the data processing inequality [44], we can bound $I(U;Z)$ as follows

$$I(U;Z) \overset{(a)}{\leq} I(X;Z) \overset{(b)}{\leq} I(X;Y)$$

where we have equality in (a), if the mapper Φ is a one-to-one function (this is not the case for many-to-one mappings, used in, e.g., some probabilistic shaping implementations [45]). In this paper, we only consider one-to-one mapping functions and thus have $I(U;Z) = I(X;Z)$. We have equality in (b) if and only if Z constitutes a sufficient statistic for X given Y [46], i.e., if X is independent of Y given Z.

While equality in (a) is obtained in most communication systems, we do not necessarily have equality in (b), especially if we employ a mismatched decoder, i.e., when the PDF $q_{Y|X}(y|x)$ assumed in the decoder does not exactly correspond to the average channel PDF $p_{Y|X}(y|x)$. Therefore, we cannot directly use $I(X;Y)$ but need to find a more accurate estimate of $I(U;Z)$ based on X and Y.

Unfortunately, in general, $p_{Y|X}$ is not known and must be estimated from the experiment. As the noise in uncompensated coherent optical fiber communication tends to be Gaussian [31], a good choice is to approximate $p_{Y|X}(y|x)$ by a Gaussian PDF, with different levels of refinement [22]. In most cases, circularly symmetric Gaussian PDFs are enough, which is what we have used in (2). To get a more accurate estimate of the

conditional channel PDF, we can also use a kernel density estimator (KDE) [47] to approximate the PDF.

As estimating the PDF $p_{Y|X}(y|x)$ is not always straightforward and because we may use a mismatched decoder with $I(U;Z) \leq I(X;Y)$, we propose to use $\underline{I}(X;Y) \leq I(X;Y)$ given in (10) as performance predictor, which originates from [42], and which we found to accurately predict $I(U;Z)$ and hence the NB-FEC performance.

In the optical communications literature, the *auxiliary channel lower bound* [48], is frequently used to estimate the MI [22] [25, Sec. III] [49, Sec. 2] [50] and which is given by

$$\begin{aligned}\underline{\underline{I}}(X;Y) &:= \mathbb{E}_{p_{X,Y}}\left\{\log_2\left(\frac{q_{Y|X}(Y|X)}{\sum_{j=1}^{M} q_{Y|X}(Y|s_j)\lambda_j}\right)\right\} \\ &\leq I(X;Y).\end{aligned} \tag{12}$$

The expectation in (12) is taken over the actual (averaged) joint channel PDF $p_{X,Y}(\cdot,\cdot)$ and $q_{Y|X}(\cdot|\cdot)$ is an *auxiliary* PDF. If $q_{Y|X}(y|x) = p_{Y|X}(y|x)$, we have $\underline{\underline{I}}(X;Y) = I(X;Y)$. Note that (12) is just a special case of (10) with $\nu = 1$ and hence

$$\underline{\underline{I}}(X;Y) \leq \underline{I}(X;Y) \leq I(X;Y),$$

where the first inequality is obvious as $\underline{\underline{I}}(X;Y)$ is recovered for $\nu = 1$ in (10) and the second inequality is shown in [42].

It is often claimed in the above-mentioned references that one should use the same $q_{Y|X}(y|x)$ as we use in the decoder (e.g., to compute the LLRs in (1)) to estimate the MI via (12). However, we found in numerical experiments that $\underline{\underline{I}}(X;Y)$ can significantly underestimate $I(U;Z)$ in many practical applications. We illustrate this discrepancy by means of an example.

Example 1: Consider the following toy example for $D = 1$ where $p_{Y|X}(y|x) = \mathcal{N}(x, \sigma_n^2)$, i.e., is Gaussian distributed with variance σ_n^2 and mean x and where $q_{Y|X}(y|x) = \mathcal{N}(x, K)$, i.e., the receiver assumes a Gaussian distribution with different variance $K \neq \sigma_n^2$. In this case, we can show that $I(X;Y) = I(U;Z)$, as we can represent $p_{Y|X}(y|x) = a(x,z)b(y)$ [46, Sec. 1.10] [46, Lem. 4.7]. The random variable Z is an $M-1$ dimensional vector with entries Z_i and realizations z_i. Let $z_i = \log\left(\frac{q_{Y|X}(y|s_i)}{q_{Y|X}(y|s_1)}\right)$. We can thus write, for $i \in \{1, \dots M\}$,

$$\begin{aligned}&q_{Y|X}(y|x) = \\ &q_{Y|X}(y|s_1)\exp\left((1-\mathbb{1}_{\{x=s_1\}})\sum_{i=2}^{M} z_{i-1}\mathbb{1}_{\{x=s_i\}}\right)\end{aligned}$$

where $\mathbb{1}_{\{\cdot\}}$ is the binary indicator function. Relating $p_{Y|X}(y|x)$ to $q_{Y|X}(y|x)$ yields

$$\begin{aligned}&\left(\sqrt{2\pi\sigma_n^2}\right)^{\frac{\sigma_n^2}{K}-1}\sqrt{\frac{\sigma_n^2}{K}}\left[p_{Y|X}(y|x)\right]^{\frac{\sigma_n^2}{K}} \\ &= q_{Y|X}(y|s_1)\exp\left((1-\mathbb{1}_{\{x=s_1\}})\sum_{i=2}^{M} z_{i-1}\mathbb{1}_{\{x=s_i\}}\right)\end{aligned}$$

which allows us to write

$$\begin{aligned}p_{Y|X}(y|x) = &\underbrace{\exp\left(\frac{K}{\sigma_n^2}(1-\mathbb{1}_{\{x=s_1\}})\sum_{i=2}^{M} z_{i-1}\mathbb{1}_{\{x=s_i\}}\right)}_{=:a(x,z)}\times \\ &\underbrace{\left(\frac{K}{\sigma_n^2}\right)^{\frac{K}{2\sigma_n^2}}\left(\sqrt{2\pi\sigma_n^2}\right)^{\frac{K}{\sigma_n^2}-1}\left[q_{Y|X}(y|s_1)\right]^{\frac{K}{\sigma_n^2}}}_{=:b(y)}\end{aligned}$$

and hence we have $I(X;Y) = I(X;Z)$. However, if we evaluate $\underline{\underline{I}}(X;Y)$ from (12) for $K \neq \sigma_n^2$, we inevitably have $\underline{\underline{I}}(X;Y) < I(X;Y)$.

If we employ for example LDPC codes with the widely used min-sum decoder or the less known linear programming decoder [51], the decoding performance does not depend on $K > 0$ used for computing the LLRs and hence, $\underline{\underline{I}}(X;Y)$ will not be an adequate performance estimate and may even largely underestimate the performance, if used as performance prediction threshold. △

We therefore propose to use the generalization (10) of (12), which we found to accurately predict $I(U;Z)$ and hence the NB-FEC performance. A convenient byproduct of using $\underline{I}(X;Y)$ is the fact that it can be used to jointly estimate the MI $I(U;Z)$ and the variance of the noise. We illustrate this application in the following example.

Example 2: For the case of uncompensated links, we know that the Gaussian PDF is a good approximation of the channel PDF [31]. However, in general, as we do not know *a priori* the variance of the noise PDF, we need to estimate it. In [25, Sec. III], it is for instance proposed to estimate the noise variance from the measurement database. Here we propose to directly use the MI estimate to obtain the noise variance. As the variance is unknown, we first fix $\sigma_n^2 = \frac{1}{2}$ in (2) and then evaluate (11) as

$$\begin{aligned}I_{\text{NB}}\Big|_{\text{awgn}} = \frac{1}{\ln(2)N_m}\sup_{\nu\geq 0}\sum_{\kappa=1}^{N_m}\Bigg(&-\nu\left\|y[\kappa]-x[\kappa]\right\|^2 - \\ &\log\left(\sum_{j=1}^{M}\lambda_j\exp\left(-\nu\left\|y[\kappa]-s_j\right\|^2\right)\right)\Bigg).\end{aligned} \tag{13}$$

After carrying out the optimization over ν (for example, as highlighted above, using the Golden section search), we immediately get an estimate of the noise variance as $\hat{\sigma}_n^2 = \frac{1}{2\hat{\nu}}$, where $\hat{\nu}$ is the ν that maximizes (13). △

C. Other Thresholds

In the remainder of this paper, the accuracy of the MI as a decoding threshold will be compared against predictions based on other performance thresholds. If BICM, as explained in Section II-B and shown in Fig. 2, is used as CM scheme, the

bit-wise GMI is a good metric [4]. The GMI is computed as

$$\text{GMI} \approx -\frac{1}{N_m}\sum_{\kappa=1}^{N_m}\log_2\left(\lambda_{\phi^{-1}(y[\kappa])}\right) - \frac{1}{N_m}\sum_{i=1}^{m}\sum_{\kappa=1}^{N_m}\log_2\left(1+\exp\left((-1)^{c_i[\kappa]}\tilde{L}_i(y[\kappa])\right)\right),$$

where $c_i[\kappa]$ is the bit at bit position i ($i \in \{1,\ldots,m\}$) mapped to symbol $x[\kappa]$ and $\tilde{L}_i(y[\kappa])$ are the *bit-wise* LLRs computed according to (4). In the practically prevalent case where all symbols are equiprobable ($\lambda_i = 1/M$), we have

$$\text{GMI} \approx m - \frac{1}{N_m}\sum_{i=1}^{m}\sum_{\kappa=1}^{N_m}\log_2\left(1+\exp\left((-1)^{c_i[\kappa]}\tilde{L}_i(y[\kappa])\right)\right).$$

We assume from now on for simplicity that all constellation symbols are equiprobable, i.e., $\lambda_i = 1/M$. Computing the GMI in a nonbinary transmission scheme necessitates the use of an extra *LLR Calculator* implementing (4), which is shown in the bottom branch of Fig. 1.

Second, we use the pre-FEC BER $\frac{1}{m}\sum_{i=1}^{m} P(\hat{B}_i \neq B_i)$, and the pre-FEC SER $P(\hat{X} \neq X)$. These quantities are schematically shown at the bottom of Fig. 1. We immediately see that only the MI is directly connected to the NB-FEC decoder, and thus is the most natural threshold choice. In particular, the transmitter in Fig. 1 uses a GF(2^m)-to-bit mapper followed by a bit-to-symbol mapper $\Phi(\boldsymbol{b}) = x$, which maps the vector of bits $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)$ to a constellation symbol $x \in \mathcal{S}$. These blocks are included only so that the GMI and pre-FEC BER can be defined (and calculated) but have no operational significance for the NB-CM system under consideration, as U can be directly mapped to X. The bit labeling used in the mapper Φ affects both the GMI and pre-FEC BER, but has no impact on the actual performance of the system. At the receiver side, additionally logarithmic likelihood ratios (LLRs) are calculated ($\tilde{\boldsymbol{L}}$), and a hard-decision on the symbols is made ($\hat{X}$), which leads to a hard-decision on the bits ($\hat{\boldsymbol{B}}$).

D. Performance Prediction for BICM-ID and Multi-Level Coding

An alternative to BICM is to use BICM with iterative demapping (BICM-ID), a concept introduced in [52], [53]. The idea is to use iterative demapping to compensate for the information loss from non-ideal BMD. BICM-ID for optical communications has been studied for instance in [54]–[56], [8, Sec. 3], [57, Sec. 3], [58, Sec. 4]. In BICM-ID, iterations between the decoder and demapper are added to a possibly already iterative FEC decoder. To keep the number of iterations low, however, one can trade FEC decoder iterations for demapper iterations. The design of BICM-ID is more complex than BICM, however, BICM-ID is expected to perform very close to a maximum likelihood (ML) sequence detector, and thus, to outperform BICM.

The MI, as described in this section, is supposedly also a good performance estimator for BICM-ID systems. However, while BICM schemes with commonly used FEC implementation behave fairly universal (see also Section II-C and V), we found that this is not the case with BICM-ID. Even small changes in the channel or the modulation format can cause severe differences in the performance of BICM-ID schemes. For example, in [59], [60], we have shown that in systems with iterative differential detection for optical systems affected by phase slips, even a change of the phase slip probability can lead to significant performance differences. In BICM-ID, generally, the FEC code has to be optimized for every modification of channel and modulation format, i.e., the universality is not guaranteed. Therefore, we suggest to always explicitly carry out decoding in BICM-ID systems, as shown in, e.g., in [61] or to use MI thresholds that have been obtained with a simulation reflecting exactly the setup of the experiment.

Recently, we have shown that the use of spatially coupled (SC) LDPC codes [35] can lead to a more universal behavior when used as FEC schemes in BICM-ID [59], [60], [62]. These results are however still preliminary and mostly based on asymptotic arguments. First simulations successfully demonstrated the improved universality of SC LDPC codes.

The same argument also applies to multi-level coding (MLC) with multi-stage decoding (MSD) [12]. This scheme is capacity-achieving and hence, the MI is a good performance estimator. However, MLC with MSD is intrinsically nonuniversal and the selection of code rates has to be adapted for every change of channel, modulation format, and bit mapping [12], which is why we also recommend either to carry out decoding or to use an MI-based threshold which has been obtained from simulations of a setup identical to the one used in the transmission experiment.

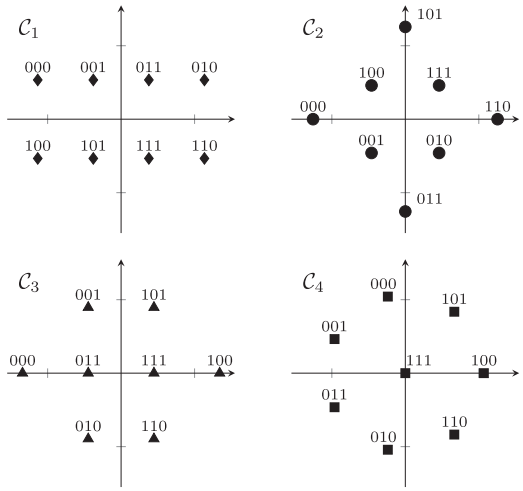

Fig. 3. Four different 8-QAM constellations used in the numerical results taken from [63]. The numbers adjacent to the constellation points give the GMI-maximizing bit labeling. The markers used for the constellation points will be subsequently used to distinguish the constellations.

IV. Experimental Verification

To experimentally verify the proposed method, we consider the four 8-QAM constellations shown in Fig. 3, where the bit-mapping that maximizes the GMI is also shown [63] [64]. For illustration purposes, we use five quasi-cyclic NB-LDPC codes with rates $R \in \{0.7, 0.75, 0.8, 0.85, 0.9\}$ (FEC

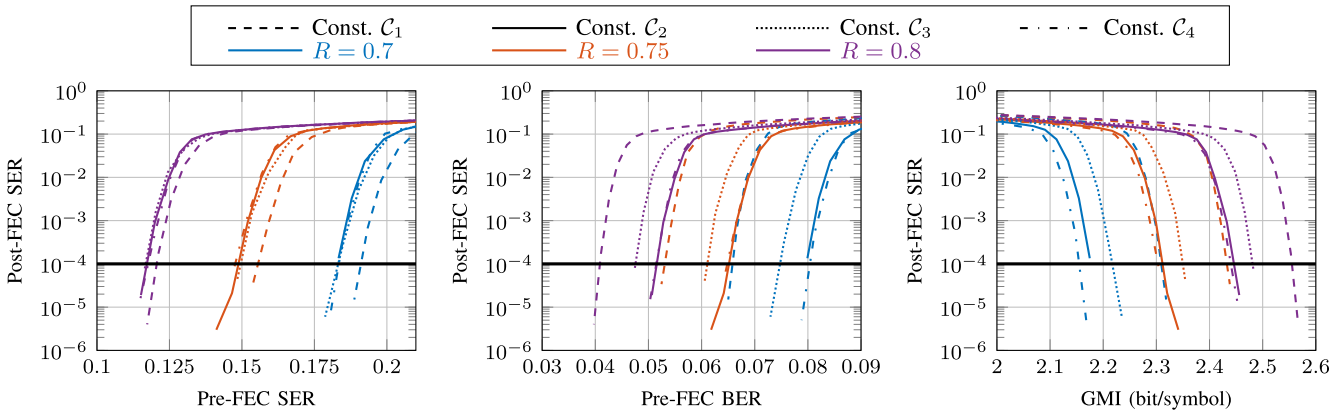

Fig. 4. Post-FEC SER as a function of three different performance metrics (pre-FEC SER, pre-FEC BER and GMI) for three NB-LDPC codes.

TABLE I
CODE PARAMETERS AND MI THRESHOLDS T_R FOR DIFFERENT CODE RATES R

Rate R	0.7	0.75	0.8	0.85	0.9
Var. degree d_v	3	3	3	3	3
Check degree d_c	10	12	15	20	30
MI threshold T_R	2.31	2.43	2.55	2.67	2.79
Normalized MI threshold T_R/m	0.77	0.81	0.85	0.89	0.93

overheads of $\approx 43, 33, 25, 18, 11\%$) defined over GF($2^3$) with regular variable node degree of $d_v = 3$ and regular check node degrees $d_c \in \{10, 12, 15, 20, 30\}$ of girth 8 ($R < 0.9$) or girth 6 ($R = 0.9$), respectively. Each code has length of around 5500, i.e., always 5500 8-QAM symbols are mapped to one LDPC codeword. The parameters of the codes are summarized in Table I. As the Galois field over which these codes are defined is rather small, the decoding complexity is relatively small as well. Decoding takes place using 15 iterations with a row-layered belief propagation decoder. These codes are conjectured to be approximately universal, i.e., their performance is expected to be independent of the actual channel (see also Section II-C).

Note that in the following we often use only a subset of constellations and code rates to keep the visualization of results simple and as we reuse previously recorded measurements. Note that the main purpose of this paper is to show that we can reuse previously recorded experimental data and evaluate the performance of NB-FEC for these experiments which is why we avoid redoing experiments.

A. AWGN Simulation Results

The performance of the five NB-LDPC codes was first tested in an AWGN channel. To this end, we first calculated the MI for the four constellations in Fig. 3. These results are shown as a function of the average symbol energy-to-noise ratio E_s/N_0 in Fig. 5 and show a clear superiority of the constellation $\mathcal{C}_4$ in terms of MI.

In Fig. 5, we also show the required E_s/N_0 for the different NB-LDPC codes to achieve a post-FEC SER of 10^{-4} and plot

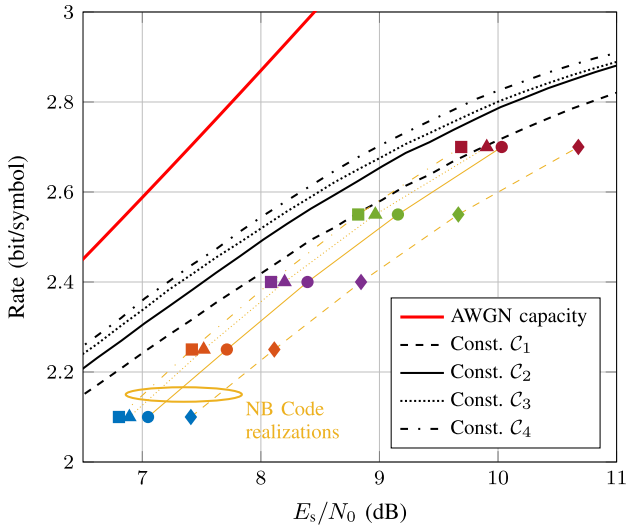

Fig. 5. MI (lines) and throughput (lines with markers) for the four 8-QAM constellations in Fig. 3 and the five NB-LDPC codes in Table I. The AWGN capacity is also shown for comparison (thick red line).

that together with the corresponding net rate, given by the number of bits per constellation symbol. The obtained results show that the NB-LDPC codes follow the MI predictions quite well, although we do observe an increasing performance gap as the code rate decreases. We attribute this loss to the nonideal code design based on the fact that we only use regular codes. Optimized irregular NB-LDPC codes [65] would be necessary for constructing better NB-LDPC codes at low rates.

In Fig. 4, we show the post-FEC SER as a function of the three performance metrics described in Section III-C for code rates $R \in \{0.7, 0.75, 0.8\}$. Changing the constellation for a given code can be interpreted as changing the nonbinary channel in Fig. 1. Additionally, in Fig. 6, we show the proposed nonbinary MI estimate $\underline{I}(X;Y)$ as performance metric for all four constellations and all five code rates. The results in Figs. 4 and 6 clearly show that only the MI can be used as a reliable threshold. In particular, for a post-FEC SER of 10^{-4} (horizontal lines in Figs. 4 and 6), the obtained MI thresholds are summarized in the third row of Table I.

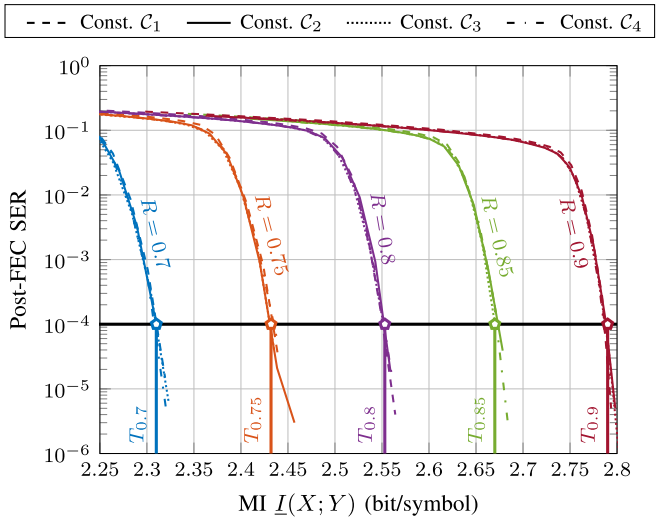

Fig. 6. Use of MI as performance metric for NB-LDPC codes.

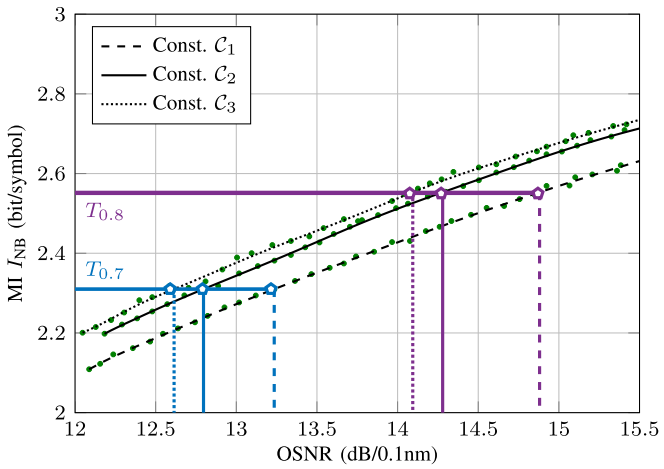

Fig. 7. Empirically obtained (green markers) and interpolated (lines) MI curves.

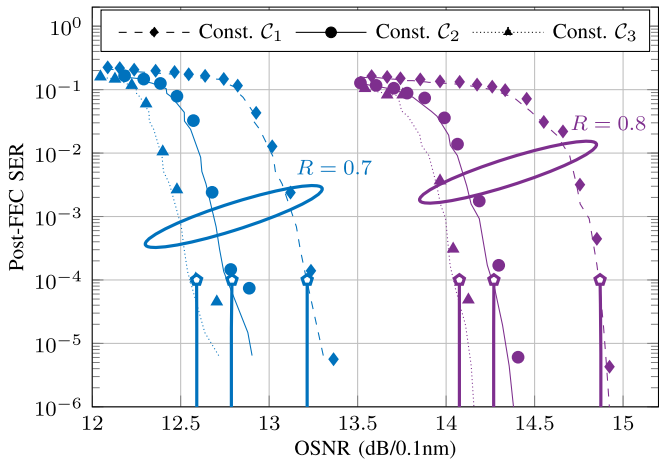

Fig. 8. Results after actual decoding with an NB-LDPC decoder with solid markers representing actual results after FEC decoding and lines representing interpolated post-FEC SER estimates taken from the estimated MI.

Instead of the MI, Fig. 4 suggests that the pre-FEC SER could also potentially serve as a performance indicator, although not as reliable as the MI. With the exception of constellation $\mathcal{C}_1$, the pre-FEC SER (which depends on the distance spectrum, i.e., the distances between constellation points) could be an indicator as well. Furthermore, for high rate codes, the pre-FEC SER becomes a better indicator. This is in line with the findings of [4], where it was shown that the GMI is the proper performance indicator for systems with BICM but for high rate codes, the pre-FEC BER can still be used with a reliability that may be good enough for some applications.

B. Back-to-Back Transmission of 8-QAM Formats

To validate the AWGN results in Fig. 4, we now consider a dual-polarization 41.6 Gbaud system. The three 8-QAM constellations of Fig. 4 were generated and tested using a high-speed DAC in a back-to-back configuration. A root-raised cosine pulse shaping (roll-off factor 0.1) signal was generated as described in [63] and two code rates ($R = 0.7$ and $R = 0.8$) were considered, giving net data rates of approximately 174 and 200 Gbit/s.

The empirical MI estimate I_{NB} as a function of the OSNR for the three constellations $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ is shown in Fig. 7, where the constellation $\mathcal{C}_3$ shows a clear superiority in terms of MI. In this figure, we also show the MI thresholds $T_{0.7} = 2.31$ and $T_{0.8} = 2.55$ from Table I. These MI thresholds are then used to determine equivalent OSNR thresholds for all three modulation formats (see vertical lines in Fig. 7). The measured data was then used to perform NB-LDPC decoding using a combination of the methods presented in [66] (scramblers) and [67] (interleavers). The obtained results are shown in Fig. 8 with solid markers. Additionally, from the estimated MI values, we interpolated the estimated post-FEC SER values using the AWGN simulations of Fig. 6, which are given by thin dashed (constellation $\mathcal{C}_1$), solid (constellation $\mathcal{C}_2$), and dotted (constellation $\mathcal{C}_4$) lines. We observe a very good agreement between the predicted post-FEC SER and actual post-FEC SER values and thus a good match between the MI thresholds obtained for the AWGN channel and the actual performance of the codes in the experiment.

C. Transmission Experiment

In order to show that the proposed method also works for a transmission over a link, we apply the method to a transmission experiment using constellations $\mathcal{C}_2$ and $\mathcal{C}_4$ over a re-circulating loop, described in detail in [64]. We recapitulate the experimental setup in the following. The transmission test-bed is depicted in Fig. 9 and consists of one narrow linewidth laser under test at 1545.72 nm, and additionally 63 loading channels spaced by 50 GHz. The output of the laser under test is sent into a PDM I/Q modulator driven by a pair of DACs operating at 65-GSamples/s. Multiple delayed-decorrelated sequences of 2^{15} bits were used to generate the multi-level drive signals. Pilot symbols and a sequence for frame synchronization are additionally inserted.

The symbol sequences are oversampled by a factor of ≈ 1.56 and pulse shaped by a root-raised cosine function with roll-off of 0.1. The load channels are separated into odd and even

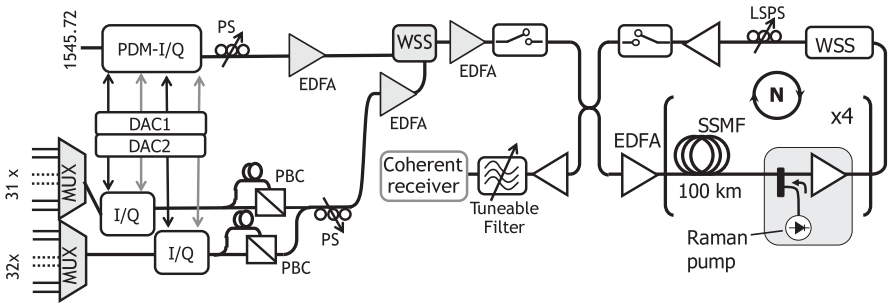

Fig. 9. WDM experimental setup with one channel under test, 63 WDM load channels, a recirculating loop consisting of four 100 km spans of SSMF and hybrid Raman-EDFA amplication.

sets of channels and modulated independently with the same constellation as the channel under test using separate I/Q modulators. Odd and even sets are then polarization multiplexed by dividing, decorrelating and recombining through a polarization beam combiner (PBC) with an approximate 10 ns delay. The test channel and the loading channels are passed into separate low-speed (<10 Hz) polarization scramblers (PS) and spectrally combined through a wavelength selective switch (WSS). The resulting multiplex is boosted through a single stage Erbium-doped fiber amplifier (EDFA) and sent into the recirculating loop. The loop consists of four 100km-long dispersion uncompensated spans of standard single-mode fiber (SSMF). Hybrid Raman-EDFA optical repeaters compensate the fiber loss. The Raman pre-amplifier is designed to provide ≈ 10 dB on-off gain. Loop synchronous polarization scrambling (LSPS) is used and power equalization is performed thanks to a 50-GHz grid WSS inserted at the end of the loop.

At the receiver side, the channel under test is selected by a tunable filter and sent into a polarization-diversity coherent mixer feeding four balanced photodiodes. Their electrical signals are sampled at 80GS/s by a real-time digital oscilloscope having a 33-GHz electrical bandwidth. For each measurement, five different sets of 20 μs are stored. The received samples are processed off-line. The DSP includes first chromatic dispersion compensation, then polarization demultiplexing by a 25-tap $T/2$ spaced butterfly equalizer with blind adaptation based on a multi-modulus algorithm.

Frequency recovery is done using 4th and 7th power periodogram for constellations $\mathcal{C}_2$ and $\mathcal{C}_4$, respectively. Phase recovery is done using the blind phase search (BPS) algorithm for both constellations. Equally-spaced test phases in the interval $[-\frac{\pi}{4}; \frac{\pi}{4})$ (constellation $\mathcal{C}_2$) or in the interval $[-\frac{\pi}{7}; \frac{\pi}{7}]$ (constellation $\mathcal{C}_4$) are used. The phase unwrapper is modified accordingly.

We consider the transmission over 8 round trips in the recirculating loop, corresponding to a distance of 3200 km. Fig. 10 shows the estimated MI I_{NB} as a function of the input power P_{in} per wavelength division multiplex (WDM) channel, see also [64, Fig. 3-a] using the Gaussian PDF $q_{\mathsf{awgn}}\big|_{D=2}$ of (2). Using a PDF estimate obtained with a KDE does not lead to noteworthy differences in the MI estimate, as predicted in [22]. Additionally, we show the MI thresholds T_R for $R \in \{0.8, 0.85, 0.9\}$. The thresholds give us the region of launch powers at which transmission is possible.

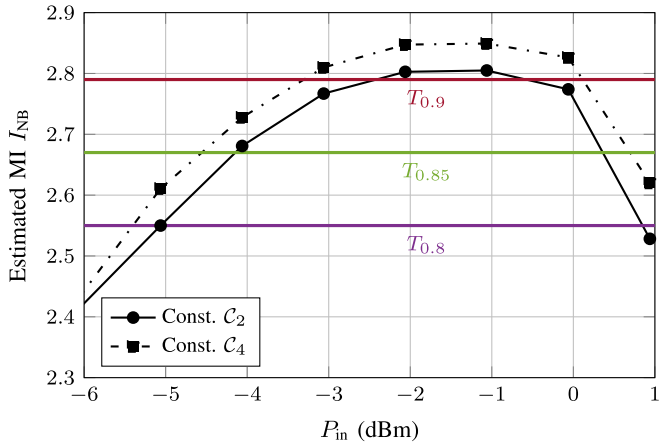

Fig. 10. Estimated mutual information for varying input power per channel for constellations $\mathcal{C}_2$ and $\mathcal{C}_4$ after transmission over 8 loops (3200 km).

To be precise, whenever the estimated MI lies above the threshold T_R, it means that successful transmission is possible, where successful is defined in the same way as for finding the threshold, i.e., with a post-FEC SER below 10^{-4}. For example, consider the red horizontal line in Fig. 10 corresponding to $T_{0.9}$. We can see that with constellation $\mathcal{C}_2$, we are just barely above the line for $P_{\text{in}} \in \{-2\,\text{dBm}, -1\,\text{dBm}\}$, which means that decoding is also only barely possible. In contrary, with constellation $\mathcal{C}_4$, we have a larger MI margin to the threshold and therefore, reliable communication is possible over a wider range of P_{in}.

In Fig. 11, we use the post-FEC SER results of Fig. 6 to estimate the post-FEC performance of the transmission system by interpolation. The interpolated curves are given by the solid (constellation $\mathcal{C}_2$) and dash-dotted (constellation $\mathcal{C}_4$) lines. Additionally, we carried out actual decoding using the LDPC codes introduced before. The post-FEC SER results after decoding are given by the solid markers in the figure. We can see that the estimates from interpolation match the actual decoding performance quite well, confirming the applicability of the proposed method.

V. Universality Revisited

In the previous sections of this paper, we have seen that MI-based thresholds can be used to accurately predict the

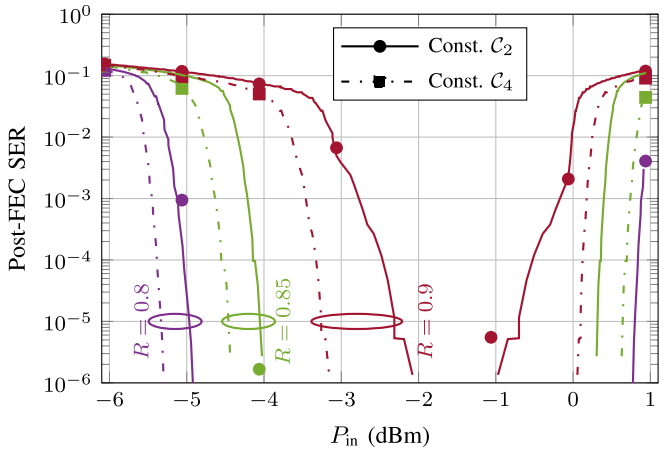

Fig. 11. Estimated post-FEC SER obtained by interpolation (curves) of the MI versus post-FEC SER obtained by actual decoding (markers) with LDPC codes of rate $R \in \{0.8, 0.85, 0.9\}$ for constellations $\mathcal{C}_2$ and $\mathcal{C}_4$ after transmission over 8 loops (3200 km).

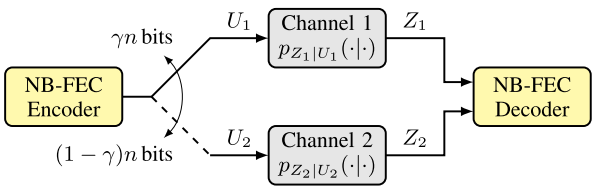

Fig. 12. Definition of universality of FEC schemes according to [34].

performance of different modulation formats with the same NB-FEC code, for which we have computed in an offline simulation an MI-threshold. However, we want to emphasize that caution must be taken: this approach assumes that the code is universal (see also Section II-C). We know from [32] that practical codes with finite block lengths are not necessarily universal.

First, we give a precise definition of the concept of universality. We can define universality of FEC schemes as in [34] with the help of Fig. 12. Consider an NB-FEC encoder that generates a codeword consisting of n symbols. We consider two different communication channels with different (memoryless) channel transition PDFs $p_{Z_1|U_1}(z_1|u_1)$ and $p_{Z_2|U_2}(z_2|u_2)$, but with *identical* MI $I_C := I(U_1; Z_1) = I(U_2; Z_2)$. A fraction γn of the symbols is transmitted over the upper channel 1, while the remaining $(1-\gamma)n$ symbols are transmitted over the lower channel 2, where $\gamma \in [0, 1]$, i.e., γ can be any real number in the unit interval, such that γn is integer. We say that a code is *universal* for channels 1 and 2 if the post-FEC SER is independent of γ. We can extend this definition to a sequence of channels and say that a code is universal if the post-FEC SER is independent of the distribution to the channels and the channels themselves.

In the previous examples of Section IV, we have not experienced any issue with universality, as the only changes we made in the channel were a change of the modulation format, but the underlying channel (AWGN or optical transmission, which can be modeled accurately as AWGN) remained fixed. In this section, we show by means of an example the impact of a more drastic change of the nonbinary channel. We now modify the channel in the AWGN simulation by adding a hard decision to the output of the optical channel. We assume that this hard decision output is based on the Euclidean distance decision metric, i.e., the output is

$$\hat{y}[\kappa] = s_{\hat{\imath}} \quad \text{with} \quad \hat{\imath} = \underset{i=1,\ldots,M}{\arg\min} \|y[\kappa] - s_i\| .$$

Although the outputs of the channel are NB hard symbols, we can still carry out soft decision decoding. In soft-decision decoding, the soft symbol demodulator calculates LLRs based on the channel statistics and the received values. Assume a memoryless optical channel and let

$$W_{j,k} := P_{\hat{Y}|X}(s_j|s_k)$$

denote the channel transition probability of receiving symbol s_j provided that symbol s_k has been sent. We can interpret this channel as a nonbinary version of the classical binary symmetric channel (BSC), often also called discrete memoryless channel (DMC). We can then compute a set of NB LLRs with

$$L_i(\hat{y}) = \ln\left(\frac{W_{\phi^{-1}(\hat{y}),i}}{W_{\phi^{-1}(\hat{y}),1}}\right) + \ln\left(\frac{\lambda_i}{\lambda_1}\right)$$

where $\phi(i) = s_i$ is the symbol mapping function. We can then use these LLRs to feed a conventional soft-decision decoder. This situation may seem at a first glance counter-intuitive, as we first make a decision and then regenerate soft-decision LLRs to use in a soft-decision NB-FEC. However, such a situation may arise when designing NB-FEC schemes for updating legacy systems that include a hard decision on symbol level which cannot be changed. The MI for this scheme is computed as

$$I_{\text{hd}} := I(X; \hat{Y}) = \sum_{i=1}^{M}\sum_{j=1}^{M} W_{j,i}\lambda_i \log_2\left(\frac{W_{j,i}}{\sum_{k=1}^{M} W_{j,k}\lambda_k}\right).$$

For illustration, we consider this scheme with the NB-LDPC codes specified in Table I and carry out a simulation over the AWGN channel with the four 8-QAM constellations shown in Fig. 3.

Figs. 13 and 14 show the post-FEC SER as a function of the pre-FEC SER after 15 LDPC decoding iterations with exactly the same decoder setup as used in Fig. 6. We can clearly see that the pre-FEC SER is again not a good performance indicator while the MI is. For comparison, we also plot in Fig. 14 the MI thresholds for the different codes from Table I. We can see that the thresholds are not as precise as previously but still reflect the actual decoding performance. We attribute this offset to the fact that the utilized LDPC codes are not exactly universal and the length of the codes is relatively small, which is an effect that has also been observed in [32]. Furthermore, we use off-the-shelf NB-LDPC codes with regular, unoptimized degree distributions. If we are allowed to increase the length of the codes and optimize the degree distributions, as highlighted for instance in [34], the performance prediction becomes more accurate again.

We hence conclude that the MI is still an accurate estimate of the NB-FEC decoding performance, even if we introduce drastic changes into the channel (like, e.g., a hard decision, going from dispersion uncompensated to dispersion compensated link, or even from coherent transmission to direct detection systems).

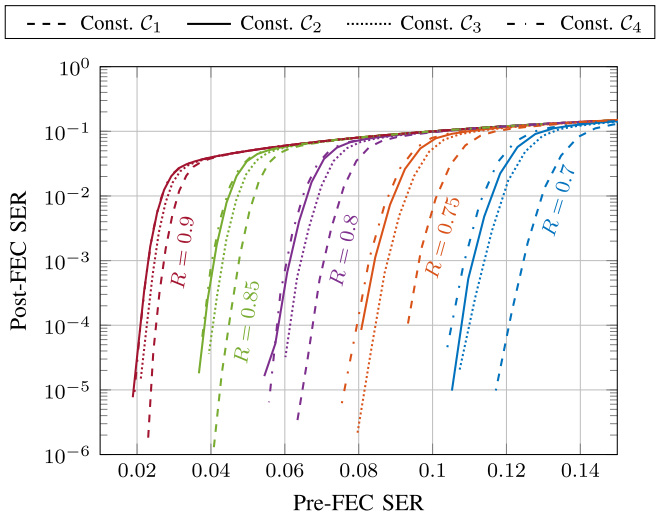

Fig. 13. Post-FEC SER as a function of the pre-FEC SER for the five LDPC codes of Table I using the four constellations of Fig. 3 after transmission over an AWGN channel with hard symbol decision at the output.

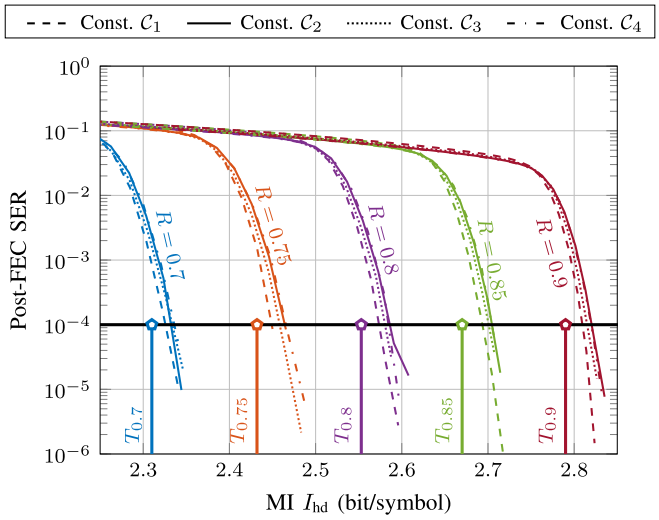

Fig. 14. Post-FEC SER as a function of the hard-decision MI I_{hd} for the five LDPC codes of Table I using the four constellations of Fig. 3 after transmission over an AWGN channel with hard symbol decision at the output.

We can improve the accuracy if the channel that is used to compute the threshold is fairly close to the channel of the system.

VI. Conclusion

Different performance metrics for coded modulation based on capacity-approaching *nonbinary* codes were compared. It was shown in simulations and experiments that an accurate predictor of the performance of these codes is the mutual information, even under severe changes of the channel. Uncoded metrics such as pre-FEC BER and pre-FEC SER were shown to fail. The GMI also fails for nonbinary codes, but still remains a good performance indicator for BICM with *binary* soft-decision FEC. We have further discussed that it is necessary that the utilized codes are universal, which is however the case for most popular FEC schemes used in optical communications.

Acknowledgements

L. Schmalen would like to thank Dr. G. Böcherer of Technical University Munich for stimulating discussions regarding mismatched decoding.

References

[1] L. Schmalen, A. Alvarado, and R. Rios-Müller, "Predicting the performance of nonbinary forward error correction in optical transmission experiments," in *Proc. Opt. Fiber Commun. Conf.*, 2016, paper M2A.2.
[2] 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.
[3] 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 *Proc. Eur. Conf. Opt. Commun.*, London, U.K., Sep. 2013, paper Th.1.C.5.
[4] 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 (Invited Paper)," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, Oct. 2015.
[5] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Corrections to "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems"," *J. Lightw. Technol.*, vol. 34, no. 2, p. 722, Jan. 2016.
[6] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.
[7] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design*. Chichester, UK: Wiley, 2015.
[8] H. Bülow, Ü. Abay, A. Schenk, and J. B. Huber, "Coded modulation of polarization- and space-multiplexed signals," in *Proc. Asia Commun. Photon. Conf. Exhib.*, Shanghai, China, Nov. 2011, pp. 1–10.
[9] 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.
[10] I. B. Djordjevic and B. Vasic, "Nonbinary LDPC codes for optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 17, no. 10, pp. 2224–2226, Oct. 2005.
[11] L. Beygi, E. Agrell, J. M. Kahn, and M. Karlsson, "Coded modulation for fiber-optic networks: Toward better tradeoff between signal processing complexity and optical transparent reach," *IEEE Signal Process. Mag.*, vol. 31, no. 2, pp. 93–103, Mar. 2014.
[12] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.
[13] A. Bisplinghoff, N. Beck, M. Ene, M. Danninger, and T. Kupfer, "Phase slip tolerant, low power multi-level coding for 64QAM with 12.9 dB NCG," in *Proc. Opt. Fiber Commun. Conf.*, Optical Society of America, 2016, paper M3A.2.
[14] G. Montorsi, "Analog digital belief propagation," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1106–1109, Jul. 2012.
[15] M. Awais, G. Masera, M. Martina, and G. Montorsi, "VLSI implementation of a non-binary decoder based on the analog digital belief propagation," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3965–3975, Aug. 2014.
[16] M. Beermann, E. Monzó, L. Schmalen, and P. Vary, "GPU accelerated belief propagation decoding of non-binary LDPC codes with parallel and sequential scheduling," *J. Signal Process. Syst.*, vol. 78, no. 1, pp. 21–34, 2015.
[17] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul., Oct. 1948.
[18] 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. 20, no. 23, pp. 1547–1549, Oct. 2011.
[19] K. Brueninghaus *et al.*, "Link performance models for system level simulations of broadband radio access systems," in *Proc. IEEE Int. Symp. Personal, Indoor Mobile Commun.*, Berlin, Germany, Sep. 2006, pp. 2306–2311.
[20] L. Wan, S. Tsai, and M. Almgren, "A fading-insensitive performance metric for a unified link quality model," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Las Vegas, NV, USA, Apr. 2006, pp. 2110–2114.

[21] E. Agrell and M. Karlsson, "Power-efficient modulation formats in coherent transmission systems," *J. Lightw. Technol.*, vol. 27, no. 22, pp. 5115–5126, Nov. 2009.

[22] T. A. Eriksson, T. Fehenberger, P. A. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments," *J. Lightw. Technol.*, vol. 34, no. 9, pp. 2256–2266, May 2016.

[23] T. A. Eriksson, P. Johannisson, E. Agrell, P. A. Andrekson, and M. Karlsson, "Biorthogonal modulation in 8 dimensions experimentally implemented as 2PPM-PS-QPSK," in *Proc. Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, Mar. 2014, paper W1A.5.

[24] T. Koike-Akino, D. S. Millar, K. Kojima, and K. Parsons, "Eight-dimensional modulation for coherent optical communications," in *Proc. Eur. Conf. Opt. Commun.*, London, U.K., Sep. 2013, paper Tu.3.C.3.

[25] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[26] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[27] T. Fehenberger, R. Maher, A. Alvarado, P. Bayvel, and N. Hanik, "Sensitivity gains by mismatched probabilistic shaping for optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 28, no. 7, pp. 786–789, Apr. 2016.

[28] E. Agrell, A. Alvarado, and F. R. Kschischang, "Implications of information theory in optical fibre communications (Invited Paper)," *Philos. Trans. Roy. Soc. A, Math. Phys. Eng. Sci.*, Vol. 374, no. 2062, Feb. 2016, Art. no. 20140438.

[29] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[30] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz),"On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[31] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 2012.

[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] 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.

[34] A. Sanaei, M. Ramezani, and M. Ardakani, "On the design of universal LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2008, pp. 802–806.

[35] L. Schmalen, V. Aref, J. Cho, D. Suikat, D. Rösener, and A. Leven, "Spatially coupled soft-decision error correction for future lightwave systems," *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1109–1116, Mar. 2015.

[36] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.

[37] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, 2009.

[38] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[39] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory (Invited Paper)," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, Aug. 2014.

[40] G. Liga, A. Alvarado, E. Agrell, and P. Bayvel, "Information rates of next-generation long-haul optical fiber systems using coded modulation," submitted to IEEE/OSA *J. Lightw. Technol.*, preprint available at https://arxiv.org/abs/1606.01689.

[41] 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.

[42] A. Ganti, A. Lapidoth, and İ. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[43] W. Cheney and D. Kincaid, *Numerical Mathematics and Computing*, 3rd ed. Pacific Grove, CA, USA: Brooks/Cole, 1994.

[44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[45] M. Yankov, D. Zibar, K. Larsen, L. Christensen, and S. Forchhammer, "Constellation shaping for fiber-optic channels with QAM and high spectral efficiency," *IEEE Photon. Technol. Lett.*, vol. 26, no. 23, pp. 2407–2410, Dec. 2014.

[46] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[47] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. CRC Press, 1986, vol. 26.

[48] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[49] 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.

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

[51] J. Feldman, M. J. Wainwright, and D. R. Karger, "Using linear programming to decode binary linear codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 954–972, Mar. 2005.

[52] 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.

[53] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping for QPSK modulation," *IEE Electron. Lett.*, vol. 34, no. 15, pp. 1459–1460, Jul. 1998.

[54] 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.

[55] 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.

[56] H. Bülow, X. Lu, L. Schmalen, A. Klekamp, and F. Buchali, "Experimental performance of 4D optimized constellation alternatives for PM-8QAM and PM-16QAM," in *Proc. Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, Mar. 2014, paper M2A.6.

[57] H. Bülow and E. Masalkina, "Coded modulation in optical communications," in *Proc. Opt. Fiber Commun. Conf.*, Los Angeles, CA, USA, Mar. 2011, paper OThO1.

[58] L. Schmalen, "Energy efficient FEC for optical transmission systems," in *Proc. Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, Mar. 2014, paper M3A.1.

[59] L. Schmalen, S. ten Brink, and A. Leven, "Spatially-coupled LDPC protograph codes for universal phase slip-tolerant differential decoding," in *Proc. Opt. Fiber Commun. Conf.*, Optical Society of America, Mar. 2015, paper Th3E.6.

[60] L. Schmalen, S. ten Brink, and A. Leven, "Advances in detection and error correction for coherent optical communications: Regular, irregular, and spatially coupled LDPC code designs," in *Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks*, X. Zhou and C. Xie, Eds., Hoboken, NJ, USA: Wiley, Mar. 2016, pp. 65–122.

[61] L. Schmalen and R. Dischler, "Experimental evaluation of coded modulation for a coherent PDM system with high spectral efficiency," in *Proc. Opt. Fiber Commun. Conf.*, 2012, paper OW1H.1.

[62] L. Schmalen and S. ten Brink, "Combining spatially coupled LDPC codes with modulation and detection," in *Proc. Int. ITG Conf. Syst., Commun. Coding*, Munich, Germany, Jan. 2013.

[63] R. Rios-Müller, J. Renaudier, L. Schmalen, and G. Charlet, "Joint coding rate and modulation format optimization for 8QAM constellations using BICM mutual information," in *Proc. Opt. Fiber Commun. Conf.*, 2015, paper W3K.4.

[64] R. Rios-Müller, J. Renaudier, P. Tran, and G. Charlet, "Experimental comparison of two 8-QAM constellations at 200 Gb/s over ultra long-Haul transmission link," in *Proc. Eur. Conf. Opt. Commun.*, Cannes, France, Sep. 2014, paper P.5.1.

[65] L. Geller and D. Burshtein, "Bounds on the belief propagation threshold of non-binary LDPC codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2639–2657, May 2016.

[66] L. Schmalen, F. Buchali, A. Leven, and S. ten Brink, "A generic tool for assessing the soft-FEC performance in optical transmission experiments," *IEEE Photon. Technol. Lett.*, vol. 24, no. 1, pp. 40–42, Jan. 2012.

[67] N. Stojanovic, Y. Zhao, D. Chang, Z. Xiao, and F. Yu, "Reusing common uncoded experimental data in performance estimation of different FEC codes," *IEEE Photon. Technol. Lett.*, vol. 25, no. 24, pp. 2494–2497, Dec. 2013.