Rate Adaptation and Reach Increase by Probabilistically Shaped 64-QAM: An Experimental Demonstration

Fred Buchali, Fabian Steiner, *Student Member, IEEE*, Georg Böcherer, *Member, IEEE*, Laurent Schmalen, *Member, IEEE*, Patrick Schulte, and Wilfried Idler

(Post-Deadline)

***Abstract*—A transmission system with adjustable data rate for single-carrier coherent optical transmission is proposed, which enables high-speed transmission close to the Shannon limit. The proposed system is based on probabilistically shaped 64-QAM modulation formats. Adjustable shaping is combined with a fixed-QAM modulation and a fixed forward-error correction code to realize a system with adjustable net data rate that can operate over a large reach range. At the transmitter, an adjustable distribution matcher performs the shaping. At the receiver, an inverse distribution matcher is used. Probabilistic shaping is implemented into a coherent optical transmission system for 64-QAM at 32 Gbaud to realize adjustable operation modes for net data rates ranging from 200 to 300 Gb/s. It is experimentally demonstrated that the optical transmission of probabilistically shaped 64-QAM signals outperforms the transmission reach of regular 16-QAM and regular 64-QAM signals by more than 40% in the transmission reach.**

***Index Terms*—Distribution matching, optical fiber communication, probabilistic amplitude shaping, rate-adaptation, spatially coupled codes.**

I. Introduction

TODAY's optical long haul transmission systems are mainly based on coherent reception technology using dual polarization quadrature amplitude modulation (QAM) and support data rates of 100 up to 400 Gb/s. Key design parameters of these systems are the spectral efficiency and the maximum transmission reach to optimally exploit all available resources, especially the fiber infrastructure. For example, a transceiver that operates on a short network segment with high signal-to-noise ratio (SNR) should achieve a high spectral efficiency to maximize the net data rate over this segment. Similarly, a transceiver operating on a long network segment (e.g., an intercontinental route) with low SNR should use either a lower order modulation format or a forward error correction (FEC) code with high overhead to ensure reliable transmission. For flexibility, today's coherent optical transceivers typically use a handful of modulation modes based on standard QAM formats. For these standard QAM formats, a gap to Shannon capacity remains which can be expressed in terms of a penalty in SNR. The standard QAM formats only offer a coarse granularity in spectral efficiency and hence also a coarse granularity in the achievable transmission reach. To overcome this lack in granularity, hybrid modulation formats have been proposed. Here, symbols carrying different modulation formats are combined; by the ratio of symbols carrying the different modulation formats the spectral efficiency can be adjusted in fine steps [2]. Even if the composition of hybrid formats [3], [4] is optimized or if orthogonal frequency-division multiplexing (OFDM) with optimized bit-loading and power-loading per subcarrier [5] is used, the gap to Shannon capacity cannot be reduced. Another degree in flexibility is supported by the variability in baud rate. Varying the baud rate enables fine variations of the bit rate, nevertheless the spectral efficiency remains unchanged.

To overcome the gap to Shannon capacity of QAM constellations, modulation formats that have a Gaussian-like shape can be used [6]–[8]. Recently, coded iterative polar modulation (IPM) [9] has been proposed for subcarrier modulation of an OFDM system to reduce the gap to Shannon capacity [10]. Theoretical shaping gains of up to 0.9 dB in SNR have been reported for higher order 256 point IPM for an extremely high spectral efficiency. The SNR required by 256 point IPM only allows for transmission over very few fiber spans of standard single mode fiber. At lower spectral efficiencies that support the targeted transmission distances for long haul systems, the shaping gain of this approach is reduced to 0.27 dB, only. The implementation of systems using IPM requires a huge effort in digital signal processing (DSP) (e.g., iterative FEC decoding and demodulation) and the granularity in spectral efficiency is coarse.

Manuscript received October 26, 2015; revised December 4, 2015; accepted December 9, 2015. Date of publication December 16, 2015; date of current version March 3, 2016. The work of G. Böcherer and P. Schulte was supported by the German Federal Ministry of Education and Research in the framework of an Alexander von Humboldt Professorship. F. Steiner was supported by the TUM–Institute for Advanced Study, financially supported by the German Excellence Initiative and the European Union Seventh Framework Program under Grant n° 291763. The work of F. Buchali and L. Schmalen was supported by the German BMBF in the scope of the CELTIC+ project SASER/SaveNet. This paper was presented in part at the 41st European Conference on Optical Communications, Valencia, Spain, September 2015 in Paper PDP3.4 [1].

F. Buchali, L. Schmalen, and W. Idler are with the Alcatel-Lucent Bell Labs, Stuttgart 70499, Germany (e-mail: first.last@alcatel-lucent.com).

F. Steiner is with the Fachgebiet Methoden der Signalverarbeitung, Technical University of Munich, Munich 80333, Germany (e-mail: fabian.steiner@tum.de).

G. Böcherer and P. Schulte are with the Institute for Communications Engineering, Technical University of Munich, Munich 80333, Germany (e-mail: georg.boecherer@tum.de; patrick.schulte@tum.de).

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.2015.2510034

64 QAM 4 different distribution, 1.a all symbols have the same probability the entropy is omehow large, if we change the probability when we increase the probabilities of the inner most symbols the entropy deacreases the entropy of x is small its a disadvantaged if we lower the entropy the capacity will decrees is there an optimal trade off? probably shaping

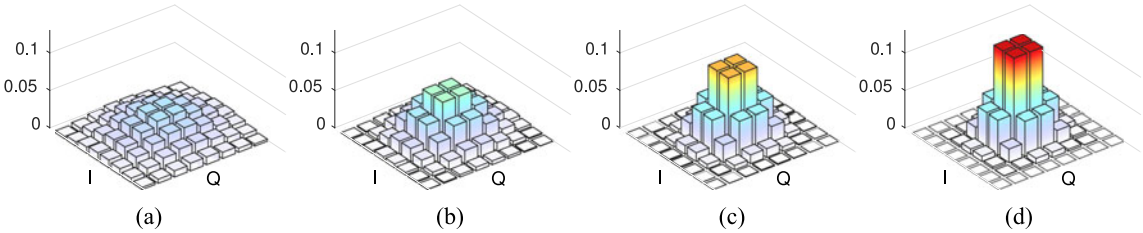

Fig. 1. Graphical illustration of the four employed probability distributions for PS-64-QAM. The bars indicate the probability of each modulation symbol. From (a) to (d), the distributions become more shaped and the entropies $\mathbb{H}(P_i)$ decrease. (a) $\mathbb{H}(P_1) = 5.73$ bits, (b) $\mathbb{H}(P_2) = 5.23$ bits, (c) $\mathbb{H}(P_3) = 4.60$ bits, (d) $\mathbb{H}(P_4) = 4.13$ bits.

Beside the characteristics of the modulation formats in the linear regime of fiber transmission, the nonlinear limitations are of high importance. Recently an improved nonlinear tolerance for shaped constellations has been reported [6]. The experimental investigation of nonlinear limits of shaped constellations is still missing. Today's transponders mostly apply pragmatic modulation formats based on bit-interleaved coded modulation (BICM). In a system supporting several modulation formats, the FEC decoder in the receiver has to decode different formats, e.g., QPSK, 16-QAM and 64-QAM. All these formats may potentially require separate FEC engines for coding and decoding, leading to an increased implementation effort.

In this paper, we propose a system that uses probabilistic constellation shaping (PS) to close the gap to Shannon capacity. The system design has an unprecedented flexibility in terms of transmission rate without increasing the system and implementation complexity. We extend our first experimental investigations [1] and experimentally verify a coded modulation scheme with rate adaptation [11] that substantially increases the transmission distance and that is flexible even though it uses a fixed FEC overhead, fixed modulation format and fixed baudrate and bandwidth. The key step is to introduce a distribution matcher (DM) [12] that generates a non-uniform modulation symbol sequence (see Fig. 1) from the data sequence. We find that the gains predicted by theory and simulations can be achieved with a practical, low-complexity system.

This paper is organized as follows: In Part 1, we describe probabilistic shaping, achievable rate estimation, rate adaptation, and FEC coding. In Part 2, we apply probabilistic shaping in an optical transmission system. In back-to-back (B2B) experiments and transmission experiments over a fiber loop, we assess the shaping gain in terms of OSNR sensitivity and in terms of maximum transmission reach. We estimate achievable rates from the measurements, which we verify by operating points where our FEC engine supports error-free transmission.

PART 1: PROBABILISTICALLY SHAPED CODED MODULATION

II. PROBABILISTIC SHAPING

We use the rate adaptive coding and modulation scheme proposed in [11], [13], which transmits different modulation symbols with different probabilities. Fig. 2 shows the high-level model. The key device is the DM [12], [14], which transforms the sequence of data bits into a sequence of non-uniformly distributed (shaped) symbols. The shaped symbols are represented

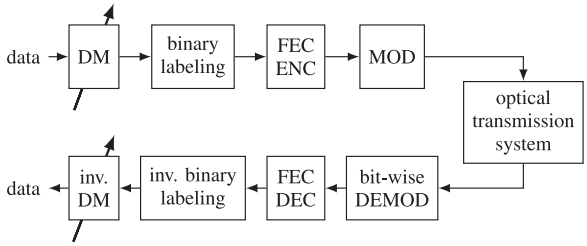

Fig. 2. System model of coding and modulation.

by binary labels and encoded by a binary FEC encoder, which is systematic to preserve the distribution of the shaped symbols. In difference to other shaping schemes (see [11, Sec. II] and references therein) shaping is done prior to FEC encoding. This reverse concatenation separates shaping and FEC completely, which results in high flexibility and low complexity. The FEC encoder output is mapped to a sequence of QAM symbols. This sequence is fed to the optical transmission system, which outputs a noisy sequence of complex QAM symbols. The demodulator uses the noisy sequence to calculate bit-wise log-likelihood ratios (LLRs), which are fed to the FEC decoder. The decoded symbols are transformed back to the data bits by an inverse DM. We next detail the different components of the scheme. For sake of clarity, we describe the scheme for a real bi-polar Amplitude Shift Keying (ASK) constellation; two consecutive ASK symbols are then mapped to the in-phase and quadrature components, respectively, of one QAM symbol.

A. Distribution Matching

The DM transforms k uniformly distributed data bits $U^k = U_1 U_2 \dots U_k$ into n amplitudes $A^n = A_1 \dots A_n$ with a desired distribution P. The DM is invertible, i.e., the data bits U^k can be recovered from the symbols A^n by applying the inverse transformation. We use Constant Composition Distribution Matching (CCDM) as proposed in [12]. The output sequences of CCDM all have the same empirical distribution. The empirical distribution of a row vector $\boldsymbol{c} = c_1 c_2 \dots c_n$ is defined as

$$\bar{P}_{\boldsymbol{c}}(a) = \frac{n_a(\boldsymbol{c})}{n} \tag{1}$$

where $n_a(\boldsymbol{c}) = |\{i : c_i = a\}|$ counts the occurrences of the symbol a in $\boldsymbol{c}$. Let $\mathcal{T}_P^n$ be the set of length n vectors with empirical distribution P, i.e.,

$$\mathcal{T}_P^n = \left\{ \boldsymbol{v} \in \mathcal{A}^n \,\middle|\, \frac{n_a(\boldsymbol{v})}{n} = P(a) \quad \forall a \in \mathcal{A} \right\} \tag{2}$$

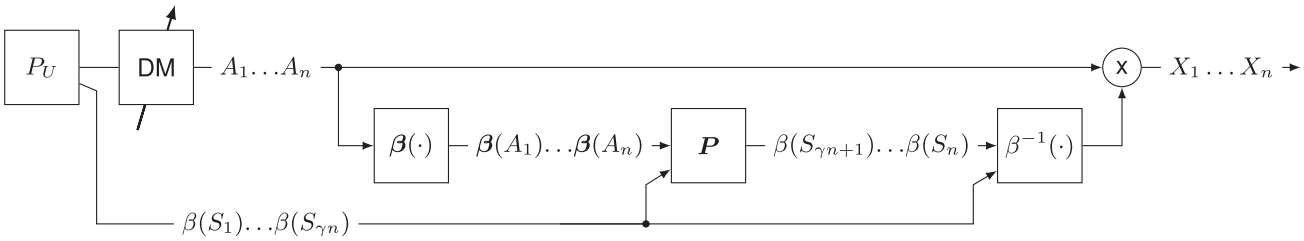

Fig. 3. PAS for code rates $(m-1+\gamma)/m$. The fraction γ of the signs is used for data, which is modelled as the output of a Bernoulli-1/2 discrete memoryless source P_U. The check bits are calculated by multiplying the binary representation $U^{\gamma n}\boldsymbol{\beta}(A^n)$ with the parity matrix $\boldsymbol{P}$ of a systematic $(m-1+\gamma)n \times mn$ binary generator matrix $[\boldsymbol{I}|\boldsymbol{P}]$, where $\boldsymbol{I}$ is the identity matrix.

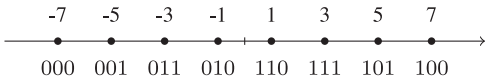

Fig. 4. The BRGC for 8-ASK. Note that the labels of the amplitudes form a 2-bit BRGC.

where $\mathcal{A}$ is the set of possible signal point amplitudes. CCDM uses vectors in $\mathcal{T}_P^n$ as codewords. There have to be at least as many codewords in $\mathcal{T}_P^n$ as there are input blocks, i.e., the input length of CCDM must not exceed $\log_2 |\mathcal{T}_P^n|$. We set the input length to $k = \lfloor \log_2 |\mathcal{T}_P^n| \rfloor$ and we define the encoding function

$$f_{\text{ccdm}} : \{0,1\}^k \to \mathcal{T}_P^n. \tag{3}$$

The actual mapping f_{ccdm} can be implemented efficiently by arithmetic coding [12], [14]. A similar scheme was proposed in [15]. For large enough n, CCDM has the following properties.

1) The marginal distributions of the output amplitudes are

$$P_{A_i} = P, \quad i = 1, 2, \ldots, n. \tag{4}$$

2) The rate is

$$\frac{k}{n} = \mathbb{H}(P) \tag{5}$$

where $\mathbb{H}(P)$ is the entropy of P in bits.

B. Probabilistic Amplitude Shaping (PAS)

1) Constellation Labeling: For each real dimension, we use an 8-ASK constellation labelled by the binary reflected Gray code (BRGC) [16], see Fig. 4. We represent the labeling by the concatenation of the sign label β and the amplitude label $\boldsymbol{\beta}$, i.e., for each 8-ASK signal point x, we have

$$\text{label}(x) = \beta(\text{sign}(x))\boldsymbol{\beta}(|x|) \tag{6}$$

where

$$\beta(-1) = 0, \ \beta(1) = 1, \tag{7}$$

$$\boldsymbol{\beta}(7) = 00, \ \boldsymbol{\beta}(5) = 01, \ \boldsymbol{\beta}(3) = 11, \ \boldsymbol{\beta}(1) = 10. \tag{8}$$

We denote the inverse functions by β^{-1} and $\boldsymbol{\beta}^{-1}$, respectively. The string $\boldsymbol{b} = b_1 b_2 b_3$ denotes the label of an 8-ASK signal point and $x_{\boldsymbol{b}}$ denotes the signal point with label $\boldsymbol{b}$.

2) Encoding and Modulation: A diagram of PAS is displayed in Fig. 3. The n amplitudes A^n generated by the DM and γn additional data bits $U^{\gamma n}$ are fed to the PAS, which computes n signs $S^n = S_1 \ldots S_n$. The first γn signs are given by $\beta^{-1}(U^{\gamma n})$. The other $(1-\gamma)n$ signs are obtained by using a rate $(m-1+\gamma)/m$ systematic binary FEC encoder to calculate $(1-\gamma)n$ check bits from the binary string $U^{\gamma n}\boldsymbol{\beta}(A^n)$. The $(1-\gamma)n$ check bits are transformed into signs by $\beta^{-1}(\cdot)$. The PAS outputs the real symbols $X_i = S_i \cdot A_i$, $i = 1, \ldots, n$. The key feature of this encoding and modulation procedure is that the amplitude distribution imposed by the DM is preserved at the PAS output.

C. Bitwise Demapping

Our digital receiver consists of a bitwise demodulator and a binary soft decoder. Suppose the PAS output X is distributed according to P and let $P_{\boldsymbol{B}}$ be the corresponding distribution of the label $\boldsymbol{B} = B_1 B_2 B_3 = \text{label}(X)$. For each received noisy ASK symbol y, the binary demodulator calculates

$$L_i = \log \frac{P_{B_i}(0)}{P_{B_i}(1)} + \log \frac{q_i(y|0)}{q_i(y|1)} \tag{9}$$

with

$$P_{B_i}(b) = \sum_{\boldsymbol{a} \in \{0,1\}^3 : a_i = b} P_{\boldsymbol{B}}(\boldsymbol{a}) \tag{10}$$

$$q_i(y|b) = \sum_{\boldsymbol{a} \in \{0,1\}^3 : a_i = b} q_{\text{awgn}}^{\text{real}}(y|x_{\boldsymbol{a}}) \frac{P_{\boldsymbol{B}}(\boldsymbol{a})}{P_{B_i}(b)} \tag{11}$$

where $q_{\text{awgn}}^{\text{real}}(\cdot|\cdot)$ is the real Gaussian channel assumed by the demapper. A bitwise demodulator followed by a binary decoder is called *bit-metric decoding* (BMD) [17, Sec. II-A1], [18].

D. Complexity

The major advantage of the proposed scheme is that its complexity is similar to the pragmatic BICM scheme, frequently used in many optical communication systems. The demapper only needs to be slightly modified taking into account the prior probabilities of the systems (see (10) and (11)). The receiver does *not* need to carry out iterations between demapper and decoder, which greatly simplifies the decoder implementation. The only additional element of the transceiver chain is the distribution (de)matcher, which, however, only adds little complexity. As discussed in [12], only a few operations (mostly comparisons and re-scaling) are required per symbol, thus keeping the complexity low.

III. Achievable Rates

In this section, we focus on rates that are *achievable* by our modulation system. An overview is provided in Table I. The

TABLE I
ACHIEVABLE RATES FROM SECTION III

Actual Channel	Auxiliary Channel	Achievable Rate	Estimate
AWGN	AWGN	$\mathbb{I}_{\text{awgn}}(P_X) = \mathbb{I}(X;Y)$	
	AWGN-BMD	$\mathbb{I}_{\text{bmd}}(P_X) = \sum_{i=1}^{m} \mathbb{I}(B_i;Y_i) - [\sum_{i=1}^{m} \mathbb{H}(B_i) - \mathbb{H}(\boldsymbol{B})]$	
Optical Channel	AWGN	$\underline{\mathbb{I}}_{\text{awgn}}(P_X)$	$\hat{\underline{\mathbb{I}}}_{\text{awgn}}(P_X)$
	AWGN-BMD	$\underline{\mathbb{I}}_{\text{bmd}}(P_X)$	$\hat{\underline{\mathbb{I}}}_{\text{bmd}}(P_X)$

capacity of the optical fiber (independent of the modulation system) is discussed, e.g., in [19], [20]. In particular, [19] estimates lower bounds on the capacity, while [20] derives an upper bound.

We start by considering a generic channel with input $X^n = X_1 X_2 \dots X_n$ and output Y^n. We use QAM inputs X_i that are independent and identically distributed according to P_X and conditioned on the input, the output has the density $p_{Y^n|X^n}$. We denote realizations of X^n and Y^n by $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. The mutual information of X^n and Y^n is given by

$$\mathbb{I}(X^n;Y^n) = \mathbb{E}\left[\log_2 \frac{p_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)}\right] \tag{12}$$

where $\mathbb{E}[\cdot]$ denotes expectation, which is taken with respect to the distributions of the involved random variables, for instance, P_X^n and $p_{Y^n|X^n}$. By [21, Ch. 5], the rate

$$\mathbb{I}(P_X) = \frac{1}{n}\mathbb{I}(X^n;Y^n) \quad \left[\frac{\text{bits}}{\text{QAM symbol}}\right] \tag{13}$$

is achievable by using some code $\mathcal{C}$ with codewords C^n that have marginal distributions $P_{C_i} = P_X$ and a decoder that uses the decoding rule

$$\hat{\boldsymbol{c}} = \underset{\boldsymbol{c}\in\mathcal{C}}{\operatorname{argmax}}\, p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{c}). \tag{14}$$

Achievable means that by choosing n large enough, the probability of erroneous decoding can be made as small as desired, as long as the transmission rate is smaller than $\mathbb{I}(P_X)$.

A. Mismatched Decoding

We do not know the exact channel law $p_{Y^n|X^n}$ of our optical channel, so we cannot calculate the achievable rate (13) and we cannot implement the decoding rule (14) either. The concept of *mismatched decoding* [22], [23] is a powerful tool that allows to derive achievable rates and decoding rules in situations when the exact channel law is not known [24]–[26]. The key insight is to replace the exact channel law $p_{Y^n|X^n}$ by an auxiliary channel $q(\cdot|\cdot)$. We consider memoryless auxiliary channels of the form

$$q^n(\boldsymbol{y}|\boldsymbol{x}) := \prod_{i=1}^{n} q(y_i|x_i). \tag{15}$$

Define now

$$\underline{\mathbb{I}}_q(P_X) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{q(Y_i|X_i)}{\sum_{x'\in\mathcal{X}} P_X(x')q(Y|x')}\right] \tag{16}$$

where the expectation is taken according to the actual distributions P_{X^n}, $p_{Y^n|X^n}$ and where $\mathcal{X}$ is the QAM constellation. Note that by [24, Sec. VI] (see also [25, Sec. IV], [26])

$$\underline{\mathbb{I}}_q(P_X) \le \mathbb{I}(P_X) \tag{17}$$

with equality, if $p_{Y^n|X^n} = q^n$, i.e., for any auxiliary channel q^n, (16) provides a lower bound of (13), furthermore, by choosing a reasonable auxiliary channel, the achievable rate (16) can be made close to its upper bound (13). We cannot calculate (16), because we do not know the exact channel law needed to carry out the expectation. Nevertheless, we can estimate (16) by replacing the probabilistic average by an empirical average, i.e., for measured sample sequences $\boldsymbol{x}, \boldsymbol{y}$, we calculate

$$\hat{\underline{\mathbb{I}}}_q(P_X) = \frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{q(y_i|x_i)}{\sum_{x'\in\mathcal{X}} P_X(x')q(y_i|x')}. \tag{18}$$

B. AWGN Auxiliary Channel

Literature [27], [28] suggests that it is reasonable to model the optical channel by a memoryless AWGN channel. To account for this, we use as auxiliary channel the I/O relation

$$Y_i = hX_i + Z_i, \quad i = 1, 2, \dots, n \tag{19}$$

where X_i is a transmitted QAM symbol, h is a scalar, and Z is circular symmetric complex Gaussian with variance σ^2. The auxiliary channel is

$$q_{\text{awgn}}(y_i|x_i) = p_Z(y_i - hx_i). \tag{20}$$

We estimate the model parameters h and σ^2 from our sample sequences $\boldsymbol{x}, \boldsymbol{y}$ via

$$\hat{h} = \frac{\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}}{\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}}, \quad \hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}} - \hat{h}^2\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}) \tag{21}$$

where $(\cdot)^{\mathrm{H}}$ is the conjugate transpose. We denote the resulting achievable rate and its estimate by $\underline{\mathbb{I}}_{\text{awgn}}(P_X)$ and $\hat{\underline{\mathbb{I}}}_{\text{awgn}}(P_X)$, respectively.

C. BMD Auxiliary Channel

The achievable rate (16) assumes symbolwise demapping, while in our implementation, we use BMD, i.e., a bitwise demapper followed by a binary decoder, see Section II-C. The effect of bitwise demapping on the achievable rate can be accounted for by using an appropriate auxiliary channel. Let $\boldsymbol{B} = B_1 B_2 \dots B_m$ be the bit labels of a QAM symbol, i.e., the concatenation of the bit labels of its real and imaginary components. An achievable rate for BMD (see [18] for a detailed discussion) is $\underline{\mathbb{I}}_{\text{bmd}}(P_X)$, where

$$q_{\text{bmd}}(y|\boldsymbol{b}) = \frac{\prod_{i=1}^{m} P_{B_i}(b_i)}{P_{\boldsymbol{B}}(\boldsymbol{b})}\prod_{i=1}^{m} q_{\text{awgn},i}(y|b_i). \tag{22}$$

The binary input auxiliary channels $q_{\text{awgn},i}(y|b_i)$ are calculated by marginalization of q_{awgn}, similar to (10) and (11). The rate estimate $\hat{\underline{\mathbb{I}}}_{\text{bmd}}$ can be calculated from sample sequences by (18).

If the actual channel $p_{Y^n|X^n}$ is memoryless and Gaussian, i.e., $p_{Y^n|X^n} = q_{\text{awgn}}^n$, by [18, Th. 2], we have $\underline{\mathbb{I}}_{\text{bmd}}(P_X) =$

$\mathbb{I}_{\text{bmd}}(P_X)$, where

$$\mathbb{I}_{\text{bmd}}(P_X) = \sum_{i=1}^{m} \mathbb{I}(B_i;Y_i) - \underbrace{\left[\sum_{i=1}^{m} \mathbb{H}(B_i) - \mathbb{H}(\boldsymbol{B})\right]}_{(\star)}. \quad (23)$$

If the bit-levels are independent, then $(\star)$ is zero and (23) recovers the rate [29, Corollary 1], [30, eq. (24)]. In particular, if P_X is equal to the uniform distribution P_U, we have

$$\mathbb{I}_{\text{bmd}}(P_U) = \sum_{i=1}^{m} \mathbb{I}(B_i;Y_i). \quad (24)$$

The right-hand side of (24) is called generalized mutual information (GMI) in [30].

IV. Rate Adaptation for 64-QAM

Let P denote the constellation distribution after modulation imposed by the DM, let $c = 5/6$ denote the code rate of the FEC code (see Section V), and let $m = 2 \cdot 3 = 6$ be the number of bit-levels of 64-QAM. By [11, Sec. IV-D], the transmission rate is

$$R = \mathbb{H}(P) - (1-c) \cdot m \left[\frac{\text{bits}}{\text{QAM symbol}}\right]. \quad (25)$$

By (25), we can transmit at different rates R by changing the distribution P and using the same FEC code. Following [31], we choose P from the family of Maxwell–Boltzmann distributions, see Fig. 1 for an illustration of the four distributions P_1, P_2, P_3, P_4 and the resulting probabilistically shaped PS-64-QAM constellations that we use in our experiment. The corresponding entropies $\mathbb{H}(P_i)$ are listed in the captions of Fig. 1.

In Fig. 5, we depict the theoretically possible shaping gains that can be obtained by using the proposed four distributions on 64-QAM compared to their uniform counterparts on 16-QAM and 64-QAM. A circularly symmetric AWGN channel is assumed. For P_1 and P_2 and 64-QAM uniform as a reference, we observe a shaping gain of 0.76 and 0.84 dB, respectively. For P_3 and P_4 and the 16-QAM uniform reference, the shaping gains are 1.55 and 0.84 dB. The large gap of the third mode is both due to the shaping as well as the beginning saturation of the 16-QAM uniform constellation at 3.6 bpQs.

V. Spatially Coupled Codes

In principle, any FEC scheme can be used in our shaping scheme described in Section II. We opt for spatially coupled LDPC codes [32] because of their excellent performance. In fact, most spatially coupled codes have the outstanding property that their estimated decoding performance is close to the performance of the underlying code ensemble under maximum likelihood decoding as the block length grows, even when a very simple, practically feasible message passing decoder is used [33]–[35]. In this paper, we use left-terminated, weakly-coupled codes with syndrome former memory $\mu = 2$, which

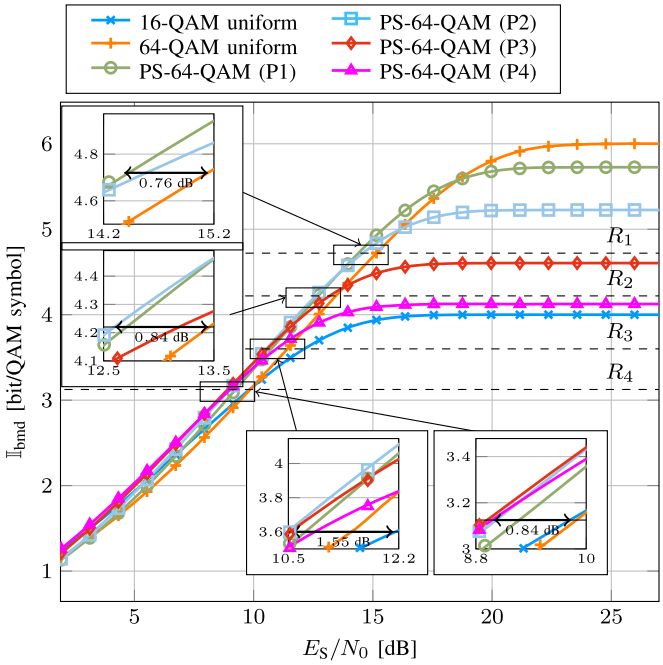

Fig. 5. Theoretic shaping gains of probabilistically shaped input distributions (P_1, P_2, P_3, P_4) compared to 16-QAM and 64-QAM uniform reference constellations on a complex AWGN channel. The transmission rates R_i are calculated by (25) using $P = P_i$, $m = 6$, and the fixed code rate $c = 5/6$.

have an infinitely extended parity-check matrix

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_0 & & & & & \\ \boldsymbol{H}_1 & \boldsymbol{H}_0 & & & & \\ \boldsymbol{H}_2 & \boldsymbol{H}_1 & \boldsymbol{H}_0 & & & \\ & \boldsymbol{H}_2 & \boldsymbol{H}_1 & \boldsymbol{H}_0 & & \\ & & \boldsymbol{H}_2 & \boldsymbol{H}_1 & \ddots & \\ & & & \boldsymbol{H}_2 & \ddots & \ddots \\ & & & & \ddots & \ddots \\ & & & & & \ddots \end{pmatrix}_{\infty}$$

with $\dim \boldsymbol{H}_i = m_{\text{sc}} \times n_{\text{sc}}$ and $n_{\text{sc}} = 6510$ for the reference codes and $n_{\text{sc}} = 7560$ for the code used for shaping. The number of rows m_{sc} of the parity-check has been chosen such that different overheads are realized. Using FPGA emulators, it has been previously shown [36]–[39] that SC-LDPC codes exhibit very low error floors and can achieve very low residual bit error rates, suitable for optical communications. The use of SC-LDPC codes hence eliminates the necessity to employ an outer cleanup code. Furthermore, SC-LDPC codes are universal and show the same error performance in different scenarios. At the receiver, we employ a windowed decoder [40] of window length $w = 13$ which carries out $I = 5$ iterations per decoding step. We use a full belief propagation decoder without any approximations to show the potential of the codes. State-of-the-art approximations like offset-corrected min-sum decoding have been shown to operate within less than 0.1 dB from the performance of the full belief propagation decoder.

TABLE II
PARAMETERS AND PERFORMANCE OF UTILIZED SPATIALLY COUPLED CODES

Code	Ref. A	Ref. B	Ref. C	Shaping
Rate	$\frac{9}{10}$	$\frac{82}{105}$	$\frac{7}{10}$	$\frac{5}{6}$
OH	$11.\bar{1}\%$	28%	43%	20%
Fraction $a_{\mathrm{v},3}$	4/5	0	0	5/6
Fraction $a_{\mathrm{v},4}$	0	1	1	0
Fraction $a_{\mathrm{v},6}$	1/5	0	0	1/6
$d_{\mathrm{c,avg}}$ (asympt.)	36	18.3	$13.\bar{3}$	21
$\frac{E_\mathrm{S}}{N_0}$ thresh. (16-QAM) [dB]	12.73	10.50	—	—
$\frac{E_\mathrm{S}}{N_0}$ thresh. (64-QAM) [dB]	—	15.78	14.16	16.84
$\mathbb{I}_{\mathrm{awgn}}$ thresh. (16-QAM) [bpQs]	3.71	3.29	—	—
$\mathbb{I}_{\mathrm{awgn}}$ thresh. (64-QAM) [bpQs]	—	4.92	4.46	5.20
$\mathbb{I}_{\mathrm{bmd}}$ thresh. (16-QAM) [bpQs]	3.70	3.28	—	—
$\mathbb{I}_{\mathrm{bmd}}$ thresh. (64-QAM) [bpQs]	—	4.90	4.43	5.18
Pre-FEC BER (16-QAM)	0.019	0.050	—	—
Pre-FEC BER (64-QAM)	—	0.052	0.078	0.038
NCG (16-QAM) [dB]	11.80	13.41	—	—
NCG (64-QAM) [dB]	—	14.34	15.49	13.57

TABLE III
GAPS TO THEORETICAL PERFORMANCE LIMITS OF UTILIZED SPATIALLY COUPLED CODES

Code	Ref. A	Ref. B	Ref. C	Shaping
Rate	$\frac{9}{10}$	$\frac{82}{105}$	$\frac{7}{10}$	$\frac{5}{6}$
OH	$11.\bar{1}\%$	28%	43%	20%
$\frac{E_\mathrm{S}}{N_0}$ gap, fix. $\mathbb{I}_{\mathrm{awgn}}$ (16-QAM) [dB]	0.644	0.671	—	—
$\frac{E_\mathrm{S}}{N_0}$ gap, fix. $\mathbb{I}_{\mathrm{awgn}}$ (64-QAM) [dB]	—	0.834	0.907	0.773
$\frac{E_\mathrm{S}}{N_0}$ gap, fix. $\mathbb{I}_{\mathrm{bmd}}$ (16-QAM) [dB]	0.595	0.627	—	—
$\frac{E_\mathrm{S}}{N_0}$ gap, fix. $\mathbb{I}_{\mathrm{bmd}}$ (64-QAM) [dB]	—	0.751	0.778	0.709
$\mathbb{I}_{\mathrm{awgn}}$ gap (16-QAM) [bpQs]	0.106	0.161	—	—
$\mathbb{I}_{\mathrm{awgn}}$ gap (64-QAM) [bpQs]	—	0.232	0.266	0.200
$\mathbb{I}_{\mathrm{bmd}}$ gap (16-QAM) [bpQs]	0.099	0.151	—	—
$\mathbb{I}_{\mathrm{bmd}}$ gap (64-QAM) [bpQs]	—	0.213	0.235	0.184

In this paper, we employ four different codes: one code of rate $c = \frac{5}{6}$ for the novel probabilistically shaped 64-QAM system and three codes that are used with the regular constellations in the reference transmission schemes. Some of the codes we use are irregular codes. We classify the codes by their (asymptotic) degree distributions where $a_{\mathrm{v},i}$ denotes the fraction of variables of degree i, i.e., the fraction of columns of $\boldsymbol{H}$ of weight i. We use check-concentrated codes of check node degrees d_c and $d_\mathrm{c}+1$. Table II lists the parameters of the codes, i.e., their design rate $c = \frac{n_\mathrm{sc}-m_\mathrm{sc}}{n_\mathrm{sc}}$, their variable node degree distributions $a_{\mathrm{v},i}$ and the average asymptotic check node degree $d_{\mathrm{c,avg}}$.

To find the decoding thresholds, we carried out simulations over an AWGN channel. As we target a residual bit error rate around 10^{-15}, we measure the decoding threshold at a post-FEC BER of 10^{-6} and then add an SNR penalty of 0.13 dB, which we have observed from previous FPGA-based simulations for similar codes [32], [38]. We further indicate in Table II the $\mathbb{I}_{\mathrm{awgn}}$ thresholds, the $\mathbb{I}_{\mathrm{bmd}}$ thresholds (as proposed in [41]) (see Section III for the definition of $\mathbb{I}_{\mathrm{awgn}}, \mathbb{I}_{\mathrm{bmd}}$) and the equivalent pre-FEC BER thresholds for the codes and the modulation format they are used with. We assume a Gray labeling of the constellation points. For completeness, we also give the *net coding gains* (NCGs) in dB taking into account uncoded transmission using the respective modulation format.

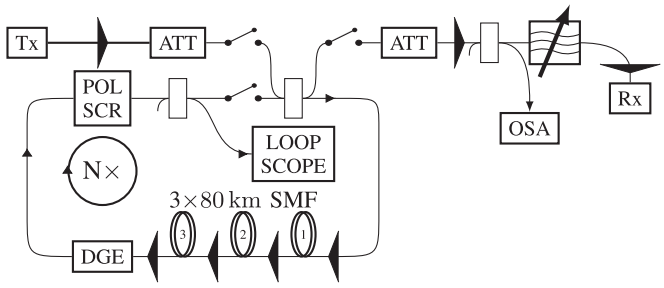

Fig. 6. Illustration of the experimental setup.

In Table III, we show the gaps of the proposed coding schemes to the rates $\mathbb{I}_{\mathrm{awgn}}, \mathbb{I}_{\mathrm{bmd}}$, which are achievable asymptotically in the blocklength. We give on the one hand the E_S/N_0 gap in decibel between the achievable rate and the decoding threshold and, on the other hand, at the decoding threshold, the vertical gap to the asymptotic rates $\mathbb{I}_{\mathrm{awgn}}, \mathbb{I}_{\mathrm{bmd}}$, expressed in bits per QAM symbols (bpQs). For computing these gaps, we used the same AWGN-simulation as above and added a 0.13 dB penalty to the threshold at 10^{-6} BER to extrapolate to a BER of 10^{-15}. We can see that all codes operate within less than 1 dB of the theoretically achievable rates. Note that all codes have approximately the same performance. In a near-term implementation using heavily constrained decoders (in terms of hardware and memory complexity, internal decoder dataflow, etc.), we have to add an additional SNR penalty, which is however identical for all codes.

PART 2: EXPERIMENTAL DEMONSTRATION

VI. EXPERIMENTAL SETUP

Optical transmission experiments have been conducted using the standard coherent transmission loop setup shown in Fig. 6. The transmitter is based on a CMOS 88 Gsamples/s quad-digital-to-analog converter (DAC) and a linear amplifier driving the dual polarization IQ-modulator. The channel under test was operated at 32 GBaud. In the transmitter (Tx) DSP, we incorporated Nyquist filtering with 0.15 roll-off factor and a pre-emphasis to compensate for the bandwidth limitations of the DAC and driver amplifier. The precalculated sequences of length 16384 are loaded into the memory of the DAC and transmitted successively. Optical modulation is carried out using a standard dual polarization IQ-modulator. A 100 kHz linewidth external cavity laser (ECL) has been used as optical carrier. In addition to the channel under test, we used 2×4 DP-QPSK load channels operated at 32 Gbaud with 4 nm guard bands to the center channel under test. These load channels increase the overall load beyond the minimum for the loop amplifiers, whereas due to the large enough guard bands, we have still single channel conditions for the center channel. The loop setup consists of three spans of standard single mode fiber (SMF, G.652) with roughly 80 km length each. The signals are amplified in single stage erbium-doped fiber amplifiers with a noise figure of 5 dB. A dynamic gain equalizer is used to compensate for differences in gain and attenuation in each round trip for the channel under test and the load channels. A slow polarization scrambler has been used to vary the alignment of the signal to the transmission fiber and hence to improve the signal statistics especially

with respect to a minor polarization dependency of the fiber, the optical amplifiers and the receiver. The fiber launch power has been adjusted either in a wide range to measure the launch power dependency and hence to find out the optimum launch power, or we set the launch power to the optimum value yielding the optimum performance at each transmission length. After a variable number of round-trips in the loop, the signals are switched to the receiver, where the channel under test is filtered out of the broad spectrum. In a dual polarization 90° hybrid, the received signal is superimposed with the local oscillator (100 kHz linewidth ECL) and O/E converted using high-bandwidth balanced photodiodes. The analog to digital conversion has been performed by a 33 GHz bandwidth real-time oscilloscope operated at 80 Gsamples/s.

Recently [1], we applied a DSP chain with blind adaptation for all unshaped and slightly shaped constellations, whereas for stronger shaped formats we found that these algorithms were not applicable without modifications. In the case of stronger shaped constellations, the quadratic shape of the constellation vanishes completely. Blind algorithms relying on the quadratic shape will consequently fail. The development of algorithms exploiting the non-quadratic constellation shape is left for future work. In [1], we implemented a data-aided receiver and applied it for those modes. We found that the performance of both DSPs was similar, but if comparing different modes in detail, the application of the same DSP is advantageous. Therefore, contrary to [1], we use in this paper a genie-aided or data-aided DSP chain for all modes. The genies' knowledge of the channel is derived from the known transmitted signal and the measured received signal. It gives an upper bound on the performance and henceforth enables the comparative study of transmission characteristics of different shaping modes. The DSP chain incorporates retiming and re-sampling to 2 bit per symbol, compensation of the chromatic dispersion, frequency offset compensation, phase estimation, and channel compensation including polarization demultiplexing. After the DSP, we carry out demodulation and soft-decision decoding as detailed in [11].

1) Offline FEC Processing

In order to allow for a practical assessment of different FEC configurations, it is beneficial to separate the FEC processing from the actual transmission experiments. Because of the offline processing, the transmitted symbols usually do not form a valid codeword and the decoding step has to be adapted to account for this. In the previous work of [42], Schmalen *et al.* assume that symmetric decoders and linear codes are used so that the all-zero codeword is always included in the codebook. The actual transmit sequence can then be thought of as being generated by modulo-2-addition of the all-zero codeword and a pseudo random scrambling sequence that is also known at the receiver so that the LLRs, which are used for soft decoding, can be descrambled again. The same idea was used in [43] by adapting the check-to-variable node message step as shown in [44] so that arbitrary symbol sequences can be decoded. Besides, the authors of [43] add additional interleavers before the decoder to artificially enlarge the set of possible transmit sequences which allows simulations to much lower BERs. Our proposed offline FEC processing borrows the idea of interleavers, but imitates the entire PAS coding and modulation scheme by incorporating the matcher and FEC encoding as well.

TABLE IV
SUMMARY OF THE REFERENCE AND PROBABILISTICALLY SHAPED MODES

	Net data rate (Gb/s)	FEC OH	Constellation
Ref1	300	28%	64-QAM
Ref2	270	43%	64-QAM
Ref3	230	11%	16-QAM
Ref4	200	28%	16-QAM
OP1	300	20%	PS-64-QAM (P_1)
OP2	270	20%	PS-64-QAM (P_2)
OP3	230	20%	PS-64-QAM (P_3)
OP4	200	20%	PS-64-QAM (P_4)

For that purpose, we transmit a number of 200000 complex 64-QAM constellation points for each of the four shaped operation modes. We regard the two-dimensional 64-QAM constellation as the Cartesian product of two 8-ASK constellations so that we store the transmitted and received symbols as a concatenation of their independent real and imaginary parts in the sets $\mathcal{X}^i_{\text{exp}}$ and $\mathcal{Y}^i_{\text{exp}}, i \in \{1,2,3,4\}$ of cardinality $N_{\text{exp}} = 400000$.

To produce a codeword $\boldsymbol{c}$ of length n_{sc} bits, we follow the PAS encoding procedure of Section II-B and let the matcher produce an amplitude sequence a^n of the right type $\mathcal{T}^n_{P_{A_i}}$ with $n = n_{\text{sc}}/m = 2170$. Since we want to operate the PAS scheme with a $c = 5/6$ code on a 8-ASK constellation with $m = 3$ bit levels, an additional number of $\gamma \cdot n = 1260$ data bits are added to the information part of the binary representation of the amplitudes. After encoding, a valid codeword $\boldsymbol{c}$ is obtained and the additional $\gamma \cdot n$ data bits along with the check bits are used as signs for the amplitudes after remapping $0 \mapsto 1, 1 \mapsto -1$. Hence, we get the sequence x^n of transmission symbols. For each $x_j, j = 1, 2, \ldots, n$, we then randomly choose a representative $x_{\text{exp},j}$ in $\mathcal{X}^i_{\text{exp}}$ such that $x_j = x_{\text{exp},j}$ and its corresponding receive symbol $y_{\text{exp},j}$ in $\mathcal{Y}^i_{\text{exp}}$. This last step corresponds to introducing interleavers to enlarge the set of transmit sequences as mentioned before. This procedure is repeated for every transmitted frame. The constructed receive sequence y^n_{exp} is used as an input for calculating the LLRs and the BER can be computed by comparing the a posteriori LLRs after BP decoding and hard-decision to the codeword $\boldsymbol{c}$.

VII. B2B TRANSMISSION RESULTS

We performed experimental characterizations of four different probabilistically shaped modes from 200 up to 300 Gb/s. For each mode we selected a corresponding reference mode based on non-shaped 16-QAM or 64-QAM with two different code rates, each. The details of all modes are summarized in Table IV.

First, we have conducted optical B2B experiments to characterize the different formats including the optical path. We received the data modulated with different formats at a variable optical SNR (OSNR). The constellations were recovered using the DSP chain as described before. Based on the noisy constellations, we calculated the quantity $\hat{\mathbb{I}}_{\text{bmd}}(P_i)$ according to (18) and (22) and plotted it versus the measured OSNR in Fig. 7. The horizontal dashed lines indicate the transmission rates for

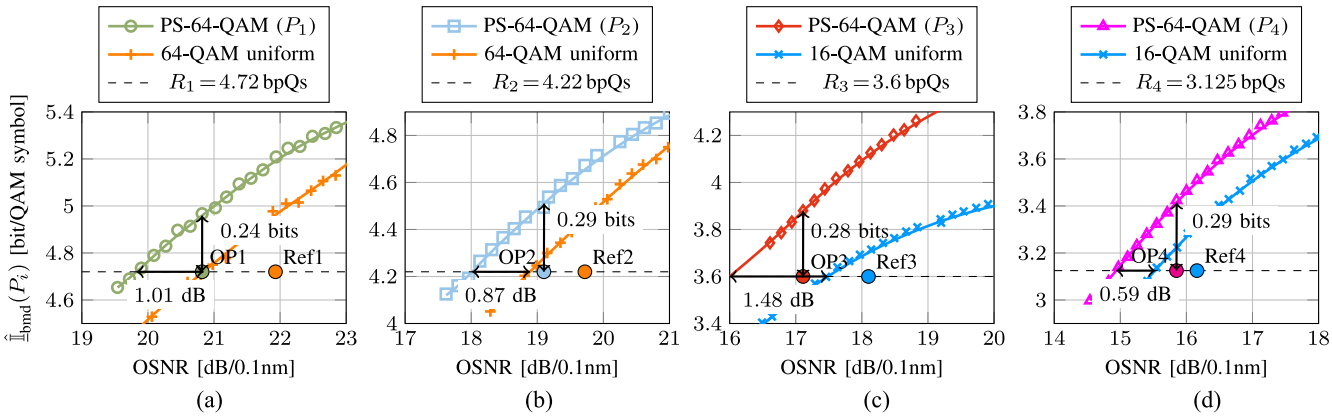

Fig. 7. Comparison of the achievable rate estimate $\hat{\mathbb{I}}_{\text{bmd}}(P)$ for shaped (P_1, P_2, P_3, P_4) and uniform scenarios in a B2B transmission. (a) Distribution P_1 (300 Gb/s), (b)Distribution P_2 (270 Gb/s), (c) Distribution P_3 (230 Gb/s), (d) Distribution P_4 (200 Gb/s).

TABLE V
GAIN OF CORRESPONDING MODES AT TRANSMISSION RATE R_i WITHOUT CONSIDERING THE FEC IMPLEMENTATION PENALTIES

Net data rate (Gb/s)	Corresponding modes	Theoretical gain (dB)	OSNR gain (dB) at R_i
300	64-QAM (28%) versus PS-64-QAM (P_1)	0.76	1.01
270	64-QAM (43%) versus PS-64-QAM (P_2)	0.84	0.87
230	16-QAM (11%) versus PS-64-QAM (P_3)	1.55	1.48
200	16-QAM (28%) versus PS-64-QAM (P_4)	0.84	0.59

TABLE VI
SUMMARY OF GAPS FROM AWGN SIMULATION AND FROM ACTUAL DECODING

	Gap at FEC thresh. from AWGN sim.		Gap at FEC thresh. from B2B experiment	
	OSNR penalty [dB]	$\mathbb{I}_{\text{bmd}}(P)$ penalty [bpQs]	OSNR penalty [dB]	$\hat{\mathbb{I}}_{\text{bmd}}(P)$ penalty [bpQs]
Ref1	1.3	0.23	0.9	0.2
Ref2	0.95	0.27	0.8	0.18
Ref3	0.7	0.11	0.6	0.1
Ref4	0.7	0.16	0.5	0.1
OP1	0.58	0.16	0.99	0.24
OP2	0.56	0.15	1.08	0.29
OP3	0.78	0.21	1.11	0.28
OP4	1.1	0.28	0.91	0.29

the different distributions as given by (25). Their crossing points with the achievable rate estimates determine the required theoretical OSNR thresholds for error free decoding without considering any implementation penalty for the FEC.

Note that the accuracy of optical B2B experiments is limited to ± 0.1 dB leading to OSNR differences of up to 0.2 dB. Still, it is possible to observe shaping gains for the four constellations in Fig. 7 as suggested by the theoretical AWGN investigations of Fig. 5. The numerical values are summarized in Table V.

Next we decoded the received data at each OSNR using the real FEC decoders and found the operating points which allow error free transmission. These are indicated by circles in Fig. 7. We also depict the rate-backoff [11, Sec. VIII], [45, Sec. IV-B], which evaluates as $\hat{\mathbb{I}}_{\text{bmd}}(P_i) - R_i$ in the given context. Using this quantity allows to compare the FEC performance of the different operation modes and is further motivated by recent work in [30] suggesting the use of the GMI as a post-FEC performance indicator. We refer to [11, Sec. VIII-C] for detailed explanations relating the backoff-criteria to the GMI. For all four distributions, we observe similar rate-backoffs of 0.24, 0.29, 0.28 and 0.29 bits/QAM symbol, which underlines the universality of the employed spatially coupled codes.

All codes have also been operated on an AWGN channel for reference. The results of both scenarios are summarized in Table VI.

VIII. TRANSMISSION RESULTS

Next we investigated the nonlinear threshold for the counterparting modes at 200 and 300 Gb/s. The optical launch power per fiber span determines the amplified spontaneous emission (ASE) noise accumulation and the accumulated nonlinear distortions. For the optimum launch power, the sum of both contributions is minimum. We varied the launch power in a wider range to determine the optimum for each mode. Fig. 8 shows $\hat{\mathbb{I}}_{\text{awgn}}(P_i)$ versus the launch power per span. For all experiments, a clear optimum has been found in the range of 0.45 to 0.5 dBm launch power. In more detail, there is a slight shift of the optimum launch power towards lower powers at an increasing number of loop round-trips. However, when connecting all maxima of the MI-versus-power plots in Figs. 8(a) and (b), a unique linear relation is observed, independent of being a uniformly distributed reference or shaped mode. For some considered transmission distances, the optimum launch power of the shaped constellation is slightly higher than for the uniform constellations. We consider the differences to be insignificant. Hence, we conclude that beside the differences in sensitivity, the modes have the same nonlinear limitations and there is no remarkable nonlinear gain for any mode.

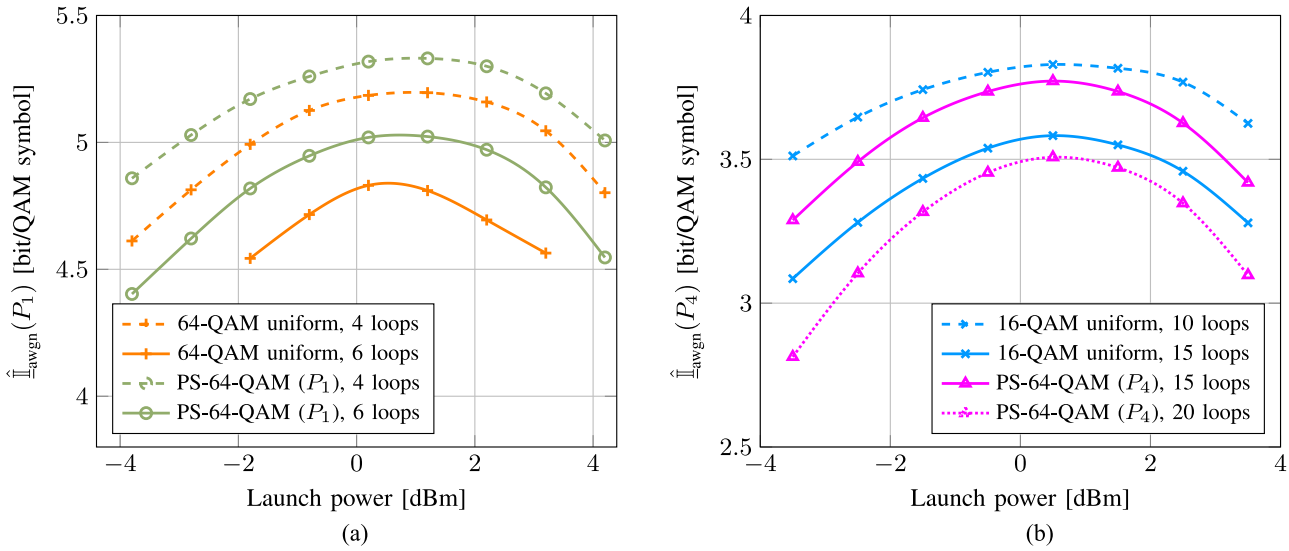

Fig. 8. Optimal launch powers for different transmission modes. (a) $R_1 = 4.72$ bpQs (300 Gb/s) (b) $R_4 = 3.125$ bpQs (200 Gb/s).

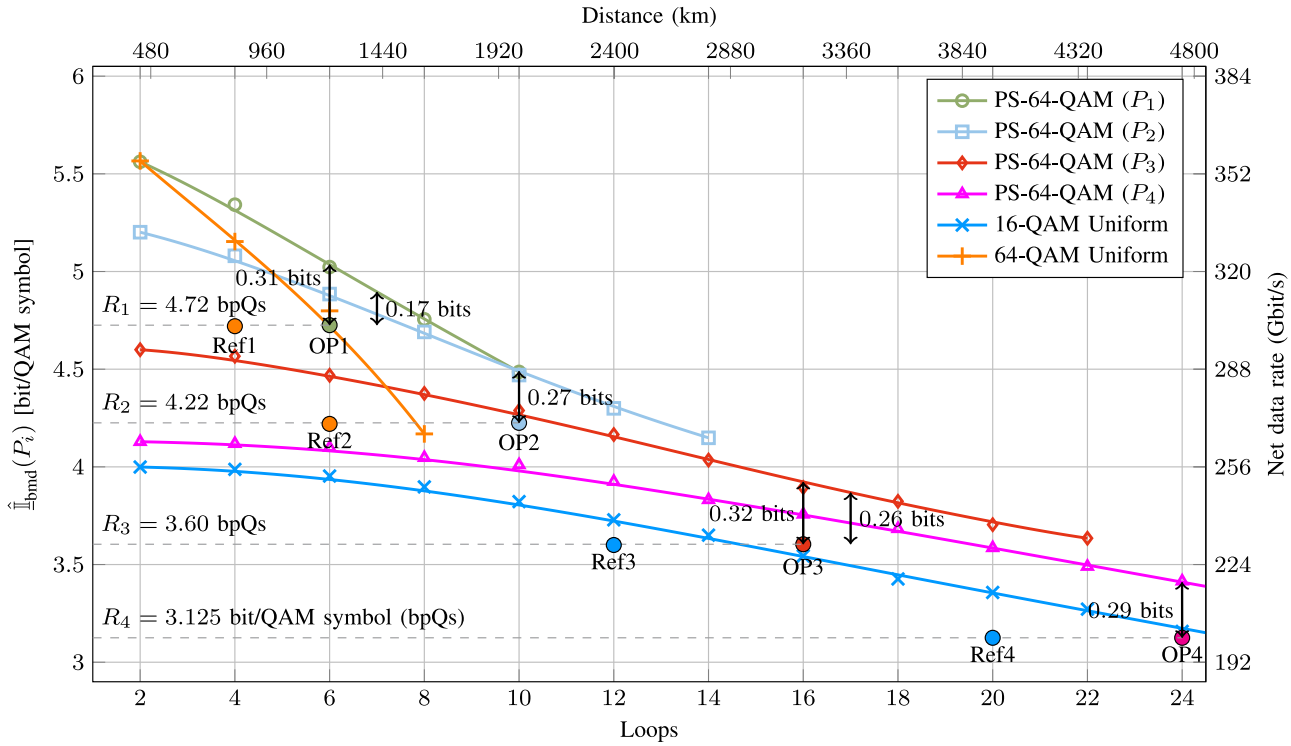

Fig. 9. Experimentally measured mutual information for the regular uniform distribution and PS-64-QAM with the four shaped distributions P_1, P_2, P_3 and P_4.

After determining the optimum launch power, we performed transmission experiments with a variable number of round-trips for all modes. The number of round-trips is plotted versus $\hat{\mathbb{I}}_{\text{bmd}}(P_i)$ in Fig. 9. Again, the dashed lines indicate the corresponding transmission rates and the circles represent the maximum number of loops such that error free decoding is still possible. Because of complexity reasons, data was collected for an even number of loops only. In the light of this, the results are consistent with those reported in Fig. 7 for the B2B experiment. Using the rate-backoff criteria as before, we obtain the backoffs 0.31, 0.27, 0.32 and 0.29 bits/QAM symbol. The conservative rate-backoffs for mode 1 and 3 in contrast to the B2B experiment are because of the coarse loop granularity. This can be seen if we interpolate and also consider an odd number of loops. In this case, the resulting rate-backoffs of 0.17 and 0.26 bit/QAM symbol would be too small to allow for successful decoding compared to the B2B case, where we have 0.24 and 0.28 bits/QAM symbol, respectively.

Compared to previous results [1], the received data is now processed by a data-aided DSP as outlined in Section VI. This improved particularly the performance of the 16-QAM reference scheme causing the shaping gain of mode 4 to decrease from six

to four loops. For all other modes, we observe improved gains, compared to [1].

IX. Conclusion

We have proposed an optical transmission system which applies probabilistic shaping to 64-QAM constellations. By having a Gaussian-like distribution over the input constellation, we are able to remove the shaping gap compared to constellations with uniform signaling. The system itself is operated at one constant modulation format and one constant FEC. System bandwidth and baud rate remain constant as well. The only adjustable block in the system is an adjustable DM, which performs the shaping of the uniform input data bits. As a result, we may adjust the net data rate with arbitrary granularity. For comparison, we use two standard modulation formats (16-QAM and 64-QAM), each combined with two different FEC engines (11%, 28% OH for 16-QAM and 28% and 43% for 64-QAM—three FEC engines in total). We demonstrated shaping gains of up to 1.4 dB OSNR derived from B2B experiments and found that shaping had no influence on the nonlinear threshold. We found a linear shaping gain improving the system by more than 40% in maximum transmission reach.

References

[1] F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Experimental demonstration of capacity increase and rate-adaptation by probabilistically shaped 64-QAM," presented at the European Conference on Optical Communication, Valencia, Spain, 2015, Paper PDP3.4.

[2] X. Zhou, L. E. Nelson, P. Magill, R. Isaac, B. Zhu, D. W. Peckham, P. I. Borel, and K. Carlson, "High spectral efficiency 400 Gb/s transmission using PDM time-domain hybrid 32-64 QAM and training-assisted carrier recovery," *J. Lightw. Technol.*, vol. 31, no. 7, pp. 999–1005, Apr. 2013.

[3] F. Buchali, L. Schmalen, K. Schuh, and W. Idler, "Optimization of time-division hybrid-modulation and its application to rate adaptive 200 Gb transmission," presented at the European Conference on Optical Communication, Cannes, France, 2014, Paper Tu.4.3.1.

[4] W. Idler, F. Buchali, L. Schmalen, K. Schuh, and H. Bülow, "Hybrid modulation formats outperforming 16QAM and 8QAM in transmission distance and filtering with cascaded WSS," presented at the Optical Fiber Communication Conference, Los Angeles, CA, USA, 2015, Paper M3G.4.

[5] L. Nadal, M. S. Moreolo, J. M. Fàbrega, A. Dochhan, H. Grießer, M. Eiselt, and J.-P. Elbers, "DMT modulation with adaptive loading for high bit rate transmission over directly detected optical channels," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 4143–4153, Nov. 2014.

[6] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2794–2798.

[7] B. Smith and F. Kschischang, "A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications," *J. Lightw. Technol.*, vol. 30, no. 13, pp. 2047–2053, Jul. 2012.

[8] 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.

[9] I. B. Djordjevic, H. G. Batshon, L. Xu, and T. Wang, "Coded polarization-multiplexed iterative polar modulation (PM-IPM) for beyond 400 Gb/s serial optical transmission," presented at the Optical Fiber Communication Conf., San Diego, CA, USA, 2010, Paper OMK2.

[10] T. Lotz, X. Liu, S. Chandrasekhar, P. Winzer, H. Haunstein, S. Randel, S. Corteselli, B. Zhu, and D. W. Peckham, "Coded PDM-OFDM transmission with shaped 256-iterative-polar-modulation achieving 11.15-b/s/Hz intrachannel spectral efficiency and 800 km reach," *J. Lightw. Technol.*, vol. 31, no. 4, pp. 538–545, Feb. 2013.

[11] 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.

[12] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 420–434, Jan. 2016.

[13] G. Böcherer, "Probabilistic signal shaping for bit-metric decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 431–435.

[14] P. Schulte, A fixed-to-fixed length distribution matcher in C/MATLAB. [Online]. Available: http://www.beam.to/ccdm

[15] T. V. Ramabadran, "A coding scheme for m-out-of-n codes," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1156–1163, Aug. 1990.

[16] F. Gray, "Pulse code communication," U. S. Patent 2 632 058, 1953.

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

[18] G. Böcherer. 2014. Achievable rates for shaped bit-metric decoding. [Online]. Available: http://arxiv.org/abs/1410.8075

[19] 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.

[20] G. Kramer, M. I. Yousefi, and F. R. Kschischang, "Upper bound on the capacity of a cascade of nonlinear and noisy channels," presented at the IEEE Information Theory Workshop, Jerusalem, Israel, Apr. 2015.

[21] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[22] T. R. Fischer, "Some remarks on the role of inaccuracy in Shannon's theory of information transmission," in *Proc. Trans. 8th Prague Conf.*, Prague, Czech Republic, 1978, pp. 211–226.

[23] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[24] 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.

[25] 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.

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

[27] L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and H. Wymeersch, "A discrete-time model for uncompensated single-channel fiber-optical links," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3440–3450, Nov. 2012.

[28] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, May 2012.

[29] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *Proc. IEEE Inform. Theory Workshop*, 2010, pp. 1517–1520.

[30] 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," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, Oct. 2015.

[31] T. Fehenberger, G. Böcherer, A. Alvarado, and N. Hanik, " LDPC Coded Modulation with Probabilistic Shaping for Optical Fiber Systems," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, Mar. 2015, Paper Th2A.23.

[32] 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.

[33] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[34] 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.

[35] M. Lentmaier, G. P. Fettweis, K. S. Zigangirov, and D. J. Costello, Jr., "Approaching capacity with asymptotically regular LDPC codes," presented at the Information Theory Applications Workshop, San Diego, CA, USA, 2009.

[36] D. Chang, F. Yu, Z. Xiao, N. Stojanovic, F. N. Hauske, Y. Cai, C. Xie, L. Li, X. Xu, and Q. Xiong, "LDPC convolutional codes using layered decoding

algorithm for high speed coherent optical transmission," presented at the Optical Fiber Communication Conf., Los Angeles, CA, USA, 2012, Paper OW1H.4.
[37] L. Schmalen, D. Suikat, D. Rosener, and A. Leven, "Evaluation of left-terminated spatially coupled LDPC codes for optical communications," presented at the Eur. Conf. Optical Communication, Cannes, France, 2014, Paper th.2.3.4.
[38] L. Schmalen, D. Suikat, D. Rösener, V. Aref, A. Leven, and S. ten Brink, "Spatially coupled codes and optical fiber communications: An ideal match?" presented at the IEEE Int. Workshop Signal Processing Advances Wireless Communication, Stockholm, Sweden, Jun. 2015.
[39] L. Schmalen, D. Suikat, D. Rösener, and V. Aref, "Next-generation forward error correction: Can we close the gap to the Shannon limit?" presented at the Asia Communications Photonics Conf., Hong Kong, China, Nov. 2015.
[40] A. R. Iyengar, M. Papaleo, P. H. Siegel, J. K. Wolf, A. Vanelli-Coralli, and G. E. Corazza, "Windowed decoding of protograph-based LDPC convolutional codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2303–2320, Apr. 2012.
[41] 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," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, Oct. 2015.
[42] 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.
[43] 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.
[44] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2141–2155, Sep. 2003.
[45] F. Steiner, G. Böcherer, and G. Liva, "Protograph-based LDPC code design for shaped bit-metric decoding," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 2, 2016.

Fred Buchali received the Diploma degree in electrical engineering and the Ph.D. degree from Humboldt-University, Berlin, Germany, in 1988 and 1991, respectively. During his Ph.D., he worked on InGaAs/InP p-i-n photodiodes for fibre-optic communication. In 1991, he joined Duisburg University, where he was involved in the research on InAlAs/InGaAs HFET and MSM-detectors. In 1992, he joined Alcatel-Lucent Germany, where he was part of the Optoelectronics Components Research Department. There, he worked on optimized pin photodiode processes for 10-Gb/s application. In 1999, he joined the optical transmission systems group, where he started research on optical and electrical impairment mitigation techniques at 10 and 40 Gb/s by analog electrical and optical techniques. During emerging coherent transmission technology, he substantially contributed to the discovery of OFDM in optics. His current research interests include coherent few carrier systems for 1Tb/s and beyond using higher order modulation formats, including the transmission aspects of the optical fiber and the digital signal processing. He was member of the OFC technical program committee and contributed continuously as reviewer for various journals. He published more than 100 journal and conference papers and holds more than 30 patent applications.

Fabian Steiner (S'14) was born in Prien am Chiemsee, Germany. He received the B.Sc. and M.Sc. degrees (with high distinction) in electrical engineering from the Technische Universität München (TUM), Munich, Germany, in 2011 and 2014, respectively. He is currently working toward the Ph.D. degree in a joint research project of the Fachgebiet Methoden der Signalverarbeitung, TUM, and University of California, Irvine, CA, USA, with the support of the TUM Institute for Advanced Study. His current research interests include coding, modulation, and multiuser massive MIMO systems. He received the third prize of the 2015 Bell Labs Prize with his proposal on probabilistic shaping for capacity achieving and rate adaptive communication.

Georg Böcherer (S'05–M'13) was born in Freiburg im Breisgau, Germany. He received the M.Sc. degree in electrical engineering and information technology from ETH Zurich, Zurich, Switzerland, in 2007, and the Ph.D. degree from the RWTH Aachen University, Aachen, Germany, in 2012. He is currently a Senior Researcher and Lecturer with the Institute for Communications Engineering, Technische Universität München, Munich, Germany. He received the E-plus award for his Ph.D. thesis and his proposal on probabilistic shaping for capacity achieving and rate adaptive communication won the third prize at the 2015 Bell Labs Prize. His current research interests include coding, modulation, probabilistic shaping for optical, wireless, and wired communications.

Laurent Schmalen received the Dipl.-Ing. degree in electrical engineering and information technology and the Ph.D. (Dr.-Ing.) degree from the RWTH Aachen University of Technology, Aachen, Germany in 2005 and 2011, respectively. He is currently a Member of the technical staff in Bell Labs' high-speed systems, signals, and processing group within the IP transport program, Bell Labs, Alcatel-Lucent. Since 2014, he also serves as Guest Lecturer at the University of Stuttgart, Germany. His research interests include forward-error correction, modulation formats, and information theory for future optical networks as well as joint source-channel coding. He received several awards including the Friedrich-Wilhelm award, the E-Plus award for his Ph.D. thesis, the best paper award of the 2010 ITG Speech Communication Conference, the 2013 best student paper award at the IEEE Signal Processing Systems workshop and the 2015 IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer Award.

Patrick Schulte was born in Dachau, Germany on April 6, 1990. He received the B.Sc. and M.Sc. degrees (with high distinction) in electrical engineering from the Technische Universität München (TUM), Munich, Germany, in 2012 and 2014, respectively. He is currently working toward the Ph.D. degree at the Institute for Communications Engineering, TUM. He is supervised by Prof. Kramer and supported by the German Aerospace Center. His current research interests include channel coding and probabilistic shaping. He received the third prize of the 2015 Bell Labs Prize with his proposal on probabilistic shaping for capacity achieving and rate-adaptive communication.

Wilfried Idler received the Diploma degree in physics from the University of Stuttgart, Stuttgart, Germany, in 1987. Between 1987 and 1996, he was with the Alcatel O/E Components Research Division in Stuttgart, involved in design and characterization of high-speed DFB lasers, tunable lasers, and wavelength converter structures, and was also involved in DFB laser technology transfers to Alcatel Optronics, France. Since 1996, he has been with the Optical Networks Division of Alcatel-Lucent Bell Labs, Stuttgart. He was contributing to the worldwide first 40-Gb/s field trials and to various Nx40 Gb/s multiterabit transmission experiments and the related conference postdeadline papers. Subsequently, he was involved in the 40-Gb/s and the 100-Gb/s product development of Alcatel-Lucent. His current research interest include coherent few carrier 1-Tb/s systems. Following his long standing experience on optical transmission systems, optical components and subsystems, he represented Alcatel-Lucent at the ITU-T study group 15 for a 10-years period. He is a Member of the technical staff in Bell Labs' high-speed systems, signals, and processing group. He was a program committee Member of OFC and regularly serves as referee for various journals. He published more than 100 journal and conference papers and holds numerous patent applications.