



## Performance Prediction Recipes for Optical Links

Downloaded from: <https://research.chalmers.se>, 2025-05-17 10:18 UTC

Citation for the original published paper (version of record):

Agrell, E., Secondini, M., Alvarado, A. et al (2021). Performance Prediction Recipes for Optical Links. *IEEE Photonics Technology Letters*, 33(18): 1034-1037.  
<http://dx.doi.org/10.1109/LPT.2021.3093790>

N.B. When citing this work, cite the original published paper.

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

# Performance Prediction Recipes for Optical Links

Erik Agrell, *Fellow, IEEE*, Marco Secondini, *Senior Member, IEEE*, Alex Alvarado, *Senior Member, IEEE*, and Tsuyoshi Yoshida, *Member, IEEE*

(Invited Tutorial)

**Abstract**—Although forward error-correction (FEC) coding is an essential part of modern fiber-optic communication systems, it is impractical to implement and evaluate FEC in transmission experiments and simulations. Therefore, it is desirable to accurately predict the end-to-end link performance including FEC from transmission data recorded without FEC. In this tutorial, we provide ready-to-implement “recipes” for such prediction techniques, which apply to arbitrary channels and require no knowledge of information or coding theory. The appropriate choice of recipe depends on properties of the FEC encoder and decoder. The covered metrics include bit error rate, symbol error rate, achievable information rate, and asymptotic information, in all cases computed using a mismatched receiver. Supplementary software implementations are available.

**Index Terms**—Achievable information rate, asymmetric information, bit error rate, mismatched decoding, optical communication, performance metrics, probabilistic shaping.

## I. INTRODUCTION

Forward error-correction (FEC) is ubiquitous in practically all deployed fiber-optical communication systems. By adding a controlled amount of overhead on top of the data payload, the receiver can correctly decode the data even in the presence of channel noise and other impairments, provided that the coding scheme and overhead are suitably chosen. A common target is a bit error rate (BER) of  $10^{-15}$  after FEC (post-FEC BER), which for a 100 Gbit/s link means about 9 bit errors per day. It may be very time-consuming and costly to reliably estimate even a single BER value in this ultralow regime. Doing it multiple times in order to test different configurations or fine-tune parameters is out of the question. Thus, system experiments and simulations are usually carried out without FEC. The topic of this brief tutorial is how to predict the performance of systems *with* FEC from performance metrics for systems *without* FEC.

The history of FEC and coded modulation for optical systems begins with *hard-decision* (HD) binary block codes and

This research was supported in part by the Swedish Research Council (VR) (grant no. 2017-03702), the Knut and Alice Wallenberg Foundation (grant no. 2018.0090), the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 757791), and National Institute of Information and Communications Technology (grant no. 20401).

E. Agrell is with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden.

M. Secondini is with the Institute of Communication, Information and Perception Technologies, Scuola Superiore Sant’Anna, Pisa, Italy, and with the National Laboratory of Photonic Networks, CNIT, Pisa, Italy.

A. Alvarado is with the Information and Communication Theory Lab, Signal Processing Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB, Eindhoven, The Netherlands.

T. Yoshida is with Information Technology R&D Center, Mitsubishi Electric Corporation, 5-1-1 Ofuna, Kamakura, 247-8501, Japan, and with Graduate School of Osaka University, 2-1 Yamadaoka, Suita, 565-0871, Japan.

binary modulation. For such codes, including the standardized (255, 239) Reed–Solomon code [1], theoretical approximations and tables accurately describe the relations between pre-FEC BER and post-FEC BER [1], [2]. The pre-FEC BER also offers good performance prediction in systems with *bit-interleaved coded modulation* [3], still under HD bit-wise decoding [4]. In systems with *soft-decision* (SD) FEC decoding [5], [6] and *probabilistic shaping* (PS) [7], [8], however, information-theoretic metrics are needed for accurate performance prediction.

A good pre-FEC performance metric is one that has a deterministic, monotonic relation with the post-FEC BER for a given FEC code, a relation that should ideally remain the same for a wide range of channels, modulation formats, and signal processing algorithms. With such a metric, alternative link configurations can be compared without implementing any FEC, and the conclusions would be valid also with FEC. Furthermore, the pre-FEC metric can be compared with a precomputed threshold to predict whether the post-FEC BER would satisfy the requirement or not.

In this paper, we give explicit expressions, or *recipes*, for predicting the post-FEC performance of transmission simulations or experiments using the most common HD and SD metrics. Importantly, they are not restricted to the additive white Gaussian noise (AWGN) channel as in some previous works. The usage and limitations of the metrics are explained, and software implementations and examples are provided [9].

## II. DATA GENERATION AND PROCESSING

The performance metrics are computed from a set of input and output data, which are generated according to the scheme in Fig. 1. The constellation consists of  $M$  symbols, represented as real-valued vectors, whose dimension  $D = 2$  in the common case of quadrature amplitude modulation (QAM). Otherwise  $D = 1$  for pulse amplitude modulation or  $D > 2$  if multiple modes (in polarization, wavelength, time, and/or space) are jointly modulated and demodulated. Each symbol is associated with an index  $1 \leq i \leq M$  and  $m = \log_2 M$  bits  $b_1(i), \dots, b_m(i)$ . An example with  $D = 2$ ,  $M = 4$ , and Gray labeling is illustrated in Fig. 1.

Throughout this paper, we are interested in FEC performance prediction for a sequence of  $N$  symbols. The data generation process starts with drawing  $N$  independent and identically distributed indices  $i_1, \dots, i_N$  from a given probability distribution  $p_j$ , which is traditionally uniform ( $p_j = 1/M$ , Sec. III–IV) and nonuniform in systems with PS (Sec. V). The indices are then mapped to the corresponding symbols  $s(i_1), \dots, s(i_N)$  in the modulator and fed to the channel,

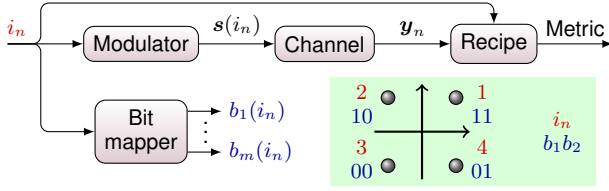


Fig. 1: Data generation, performance metric computation, and a simple bit mapping example (4-QAM).

which produces the output symbols  $\mathbf{y}_1, \dots, \mathbf{y}_N$ . In general, the channel represents either a physical channel (as in a lab experiment) or its numerical emulation (as in a computer simulation) and must include all relevant analog and digital operations (e.g., filtering, D/A and A/D conversion, and synchronization, but no FEC encoding/decoding) performed at the transmitter and at the receiver.

As discussed in the next sections, the relevant performance metric for the considered system depends on the decoding strategy (SD, HD, bit-wise, symbol-wise, ...) employed at the receiver. In an optimal receiver, the decoding strategy should also account for the exact input–output relation of the channel, which is nonlinear and bursty [10]. Since this exact relation is generally unknown, a *mismatched* receiver—designed according to a simplified (auxiliary) channel model—is usually employed. Typically, the receiver is designed by assuming that the output samples are simply corrupted by AWGN with a variance per dimension equal to

$$\sigma^2 = \frac{1}{DN} \sum_{n=1}^N \|\mathbf{y}_n - \mathbf{s}(i_n)\|^2. \quad (1)$$

We emphasize that this is not the actual behavior of the channel and is just the implicit assumption made, for instance, when using a Euclidean distance metric in the decoder. In this case, it is fundamental that the channel includes all the processing blocks that are needed to minimize the mismatch between the true channel and the adopted model, such as trivial scaling, rotation, and delay as well as symbol interleaving to disperse error bursts and more refined algorithms for clock recovery, carrier recovery, and channel equalization. The recipes in the following sections provide accurate performance predictions for matched as well as mismatched receivers [4], [11], [12].

### III. RECIPES WITH HD FEC

Hard decisions means that the receiver takes a tentative decision  $\hat{i}_n$  on the transmitted data  $i_n$  corresponding to each received vector  $\mathbf{y}_n$ . This data  $\hat{i}_n$  is then fed into the HD FEC decoder for possible error correction. The most common HD rule is the minimum Euclidean distance

$$\hat{i}_n = \arg \min_{j \in \{1, 2, \dots, M\}} \|\mathbf{y}_n - \mathbf{s}(j)\|^2, \quad (2)$$

which implicitly assumes a mismatched receiver optimized for the memoryless AWGN channel.

In HD systems, two standard pre-FEC metrics are the *symbol error rate* (SER)  $P_s$  and BER  $P_b$ , which simply count errors between the transmitted and received data:

$$P_s = \frac{1}{N} \sum_{n=1}^N \delta(i_n \neq \hat{i}_n), \quad (3)$$

$$P_b = \frac{1}{mN} \sum_{n=1}^N \sum_{k=1}^m \delta(b_k(i_n) \neq b_k(\hat{i}_n)). \quad (4)$$

The function  $\delta(\cdot)$  is 1 if the argument is true and 0 otherwise.

The SER is, albeit conceptually very simple, not very useful in practice. It is relevant for nonbinary coding using a coding alphabet of the same size as the constellation and the Hamming distance as the FEC decoding metric. AIRs for such systems are discussed in [11], [13]. However, nonbinary coded modulation is not very common in practice, since its performance gains over BICM are typically modest compared with the significantly increased complexity [14, Sec. 7.6.4].

The BER is a relevant performance metric for BICM with binary FEC and HD decoding, which is a very common setup in deployed systems. If the BER (4) is below a certain threshold, which varies depending on the selected code, then a satisfactory post-FEC performance is guaranteed, provided that sufficiently deep bit-interleaving is applied. Such BER thresholds have been tabulated for many binary codes [10], [14] of various code rates  $R_c$ . BER results and BER thresholds are often equivalently presented in terms of the *Q factor*  $Q_{\text{hard}} = \sqrt{2} \operatorname{erfc}^{-1}(2P_b)$ , which gives the theoretical signal-to-noise ratio (SNR) of a binary-input AWGN channel to achieve the same BER  $P_b$ .

The AIR using a selected code for which the BER condition is met equals  $mR_c$  bit/symbol. If on the other hand ideal FEC coding is assumed, then the AIR with HD BICM is

$$\text{AIR}_b^{\text{HD}} = m(1 - H_2(P_b)) \geq mR_c, \quad (5)$$

where  $H_2$  is the binary entropy function [11, Eq. (25)].

### IV. RECIPES WITH SD FEC

In this section, we discuss soft-decision FEC, where the receiver does not take a decision on the transmitted index (as in (2)), but instead, it uses *soft information* (reliability information) available in the received sequence  $\mathbf{y}_n$ . Here we consider two AIRs, which are valid for systems with symbol-wise and bit-wise receivers, resp., namely [15, Eq. (52)–(53)]

$$\text{AIR}_s = m - \frac{1}{N} \sum_{n=1}^N \left\{ \log_2 \sum_{j=1}^M q(\mathbf{y}_n, \mathbf{s}(j)) - \log_2 q(\mathbf{y}_n, \mathbf{s}(i_n)) \right\}, \quad (6)$$

$$\text{AIR}_b = m - \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^m \left\{ \log_2 \sum_{j=1}^M q(\mathbf{y}_n, \mathbf{s}(j)) - \log_2 \sum_{j=1}^M \delta(b_k(j) = b_k(i_n)) q(\mathbf{y}_n, \mathbf{s}(j)) \right\}, \quad (7)$$

where

$$q(\mathbf{y}, \mathbf{s}) = \exp(-\|\mathbf{y} - \mathbf{s}\|^2 / (2\sigma^2)) \quad (8)$$

and  $\sigma^2$  is given by (1).

When compared to the SER and BER recipes in (3) and (4), the AIRs in (6) and (7) only require one additional summation, and thus, their computation is still very simple. We emphasize again that the choice of  $q(\mathbf{y}, \mathbf{s})$  in (8) is arbitrary and might not correspond to the actual behavior of the channel under consideration. More details are discussed in Sec. VII.

Like the SER,  $\text{AIR}_s$  is less popular because it applies to systems based on nonbinary codes [4], and in practice binary codes are preferred.  $\text{AIR}_b$  is a relevant performance metric for BICM with binary FEC and SD decoding. Such systems are popular in practice because of the ease of implementation of binary codes and the improved performance over HD-FEC.

$\text{AIR}_b$  is an AIR with BICM and ideal binary SD-FEC. With a practical FEC, the actual rate is  $mR_c \leq \text{AIR}_b$ , while  $\text{AIR}_b$  can be used to predict the post-FEC BER. When  $\text{AIR}_b/m$  is below a certain threshold, it can be claimed that a certain post-FEC BER will be obtained at a given coding rate  $R_c$ . The exact value of the threshold depends on the specific code under consideration. Thresholds for  $\text{AIR}_b/m$  for a concatenated FEC scheme with various inner SD-FEC codes and an outer HD-FEC code (not simulated) were first published in [6, Table III]. Recently, thresholds for SD-FEC for an approximated post-FEC BER of  $10^{-15}$  have been reported in [14, Table 7.5]. An  $\text{AIR}_b$  can be converted to a Q factor as  $Q_{\text{soft}} = \sqrt{J^{-1}(\text{AIR}_b/m)/2}$ , where  $J$  is defined in [16, App.]. This is the SD analogy of  $Q_{\text{hard}}$  in Sec. III.

The metrics discussed in this section are applicable to soft-decision FEC where the bits and symbols are equally likely. The soft information passed to the decoder is typically represented using  $L$ -values (or log-likelihood ratios). These  $L$ -values can also consider systems where bits or symbols are transmitted with different probabilities, which is the main focus of the next section.

## V. RECIPES WITH PROBABILISTIC SHAPING

The most popular PS scheme in recent years is probabilistic amplitude shaping, where the shaping operation, often realized using *distribution matching* (DM), is placed outside the FEC coding operation. The DM encodes on average  $R_p$  bits onto each shaped symbol, where  $R_p$  is less than the symbol entropy  $H_s = -\sum_{j=1}^M p_j \log_2 p_j$  [18]. This system uses binary FEC and memoryless bit-wise decoding. When employing SD FEC, bit-wise  $L$ -values

$$L_{n,k} = \ln \frac{\sum_{j=1}^M \delta(b_k(j) = 0) p_j q(\mathbf{y}_n, \mathbf{s}(j))}{\sum_{j=1}^M \delta(b_k(j) = 1) p_j q(\mathbf{y}_n, \mathbf{s}(j))}, \quad (9)$$

with  $q$  given in (8), are fed into the FEC decoder for every received vector  $\mathbf{y}_n$ .

Several performance metrics for the post-FEC BER of systems with PS and SD FEC have been proposed [7], [8], [17], [18]. In this short tutorial, we give a recipe for one of them, the *asymmetric information* (ASI). To estimate the ASI, we first estimate the distribution of the *asymmetric L-values*  $L_{n,k}^a = (-1)^{b_k(i_n)} L_{n,k}$  via a histogram. For a given number of bins  $B$  with separation  $\Delta$ , we define a quantization function

$T(\ell) = \arg \min_j |\ell - \ell_j|$  where  $\ell_j = (2j - 1 - B)\Delta$  for  $j = 1, \dots, B$ . Using the probability estimates

$$\Lambda_j = \frac{1}{mN} \left| \left\{ T(L_{n,k}^a) = \ell_j : 1 \leq n \leq N, 1 \leq k \leq m \right\} \right| \quad (10)$$

for  $j = 1, \dots, B$ , the ASI estimate is

$$\text{ASI} = \sum_{j=1, \Lambda_j \neq 0}^B \Lambda_j \log_2 \left( \frac{2\Lambda_j}{\Lambda_j + \Lambda_{B+1-j}} \right). \quad (11)$$

For sufficiently high  $N$  and suitable choices of  $B$  and  $\Delta$ , the estimate (10)–(11) approaches the true ASI [8, Eq. (35)]. We suggest  $B = 32$  and  $\Delta = 1$  as default parameters.<sup>1</sup>

The same thresholds as for  $\text{AIR}_b/m$  in Sec. IV are valid for ASI, because the ASI is an achievable binary FEC code rate for uniform and PS signals and can be used to predict the post-FEC BER [17]. Similarly to the case in Sec. IV, an ASI can be converted into a Q factor as  $Q_{\text{soft}} = \sqrt{J^{-1}(\text{ASI})/2}$ . An AIR with BICM and ideal binary SD-FEC and DM is

$$\text{AIR}_b^{\text{PS}} = H_s - (1 - \text{ASI})m. \quad (12)$$

With practical (nonideal) FEC and DM, the net data rate is  $R_p - (1 - R_c)m \leq \text{AIR}_b^{\text{PS}}$ .

A suitable metric for HD PS systems with bit-wise decoding is the pre-FEC BER [8, Eq. (30)], which can be numerically estimated as

$$P_b^{\text{PS}} = \frac{1}{mN} \sum_{n=1}^N \sum_{k=1}^m \delta(L_{n,k}^a \leq 0). \quad (13)$$

It determines the maximum AIR of such systems [19, Sec. II]. In the special case of  $p_j = 1/M$ ,  $P_b^{\text{PS}}$  is similar to (4) for many channels, especially at high SNR, but not identical, since (13) applies bit-wise decisions [20] and (4) symbol-wise.

## VI. EXAMPLES

Here we consider a memoryless channel with both phase noise and AWGN, defined by  $\mathbf{y}_n = \mathbf{R}(\theta_n)\mathbf{s}(i_n) + \mathbf{Z}_n$ , which has been considered in many previous works [21]. Here  $D = 2$ , the matrix  $\mathbf{R}(\theta_n)$  denotes a rotation by a random angle  $\theta_n$ , which is zero-mean Gaussian with variance  $\sigma_\theta^2$ , and  $\mathbf{Z}_n$  is independent AWGN with variance  $\sigma_z^2$  in each dimension.

We use  $N = 10^6$  samples to obtain reasonably accurate estimates and  $\sigma_\theta^2 = 0.01$  to make the impact of phase noise clearly visible in the considered SNR range. We study uniform 64-QAM (blue,  $H_s = 6$  bit/symbol) and PS 256-QAM (green,  $H_s = 6.3$  bit/symbol), both with Gray bit mapping. Fig. 2 (top) shows the SER and BER estimates as functions of the AWGN SNR, defined as the signal variance (energy) per dimension divided by  $\sigma_z^2$ . These results show that even at high SNRs,  $P_b$  and  $P_s$  reach error floors. This is due to the fact we use (2), which is a suboptimal decision rule for this channel.  $P_b^{\text{PS}}$  with the mismatched  $q(\mathbf{y}, \mathbf{s})$  in (8) behaves similarly.

Fig. 2 (bottom) shows that the AIR  $\text{AIR}_b^{\text{HD}}$  in (5) is considerably lower than all other AIRs, which is due to the HD-FEC assumption. Fig. 2 (bottom) shows that  $\text{AIR}_s \approx \text{AIR}_b$  for

<sup>1</sup>Larger or smaller  $\Delta$  may be beneficial at very high or low SNRs, resp.

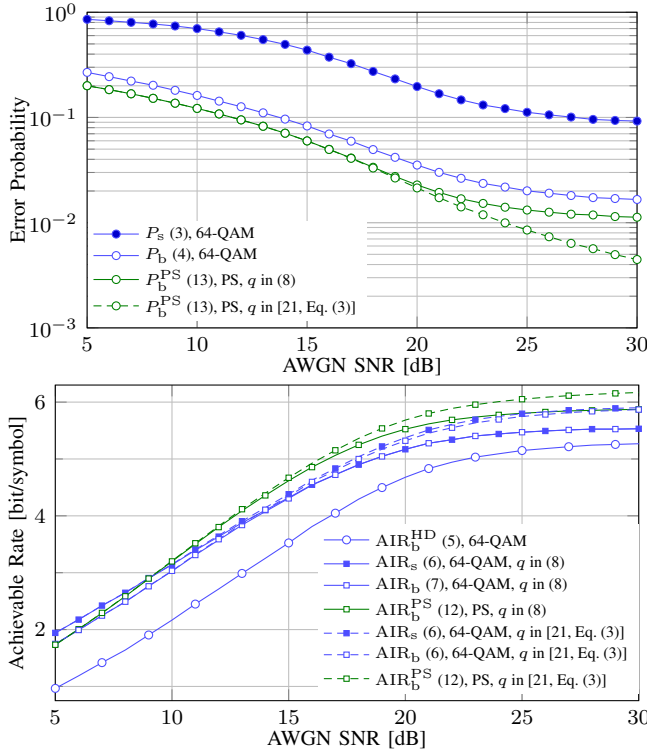


Fig. 2: Error probabilities (top) and AIRs (bottom).

AIRs above 4 bit/symbol, due to the use of a Gray labeling. Similarly to Fig. 2 (top),  $AIR_s$ ,  $AIR_b$ , and  $AIR_b^{PS}$  saturate at high SNRs, which in this case is due to the use of the mismatched  $q(\mathbf{y}, \mathbf{s})$  in (8). To improve these AIRs, we use [21, Eq. (3)]<sup>2</sup> which is a good approximation for the channel under consideration. Here both  $\sigma_\theta^2$  and  $\sigma_z^2$  are assumed to be perfectly estimated. This  $q(\mathbf{y}, \mathbf{s})$  is better matched to the channel, which results in lower  $P_b^{PS}$  and higher AIRs in Fig. 2 (dashed).

## VII. EXTENSIONS AND CONCLUSIONS

Many variants of the presented metrics exist. The recipes (3)–(7), (11), and (13) give valid performance metrics for arbitrary  $\hat{i}$  and arbitrary positive  $q$ , not only (2) and (8), as exemplified in Sec. VI. These metrics are achievable by receivers that use the same  $\hat{i}$  and  $q$ . Improved performance is often attained by optimization over various parameters. For example,  $\sigma^2$  in (8) can be optimized, which is analogous to optimization over the exponent  $s$  in [3, Eq. (4.34)], [4, Example 2]. Other options include, e.g., to let  $\sigma^2 = \sigma_i^2$  depend on the symbol index  $i$ , to define  $q$  according to a correlated, possibly multidimensional, Gaussian distribution [22], and to capture channel memory via hidden Markov models [12].

To conclude, the presented practical recipes are related to well-known information-theoretic quantities. If (6) is optimized over all positive functions  $q$ , which for general channels is possible only in theory, then the resulting AIR approaches the memoryless *mutual information* as  $N \rightarrow \infty$ . If (7) is similarly optimized over  $q$ , then the *generalized mutual*

*information* for BICM [3, Eq. (4.53)] is obtained. If the mutual information is further optimized over all possible constellations  $\mathbf{s}(i)$  and/or distributions  $p_j$  while  $M \rightarrow \infty$ , which corresponds to optimizing over all possible input distributions, then the *channel capacity* is obtained. Therefore, the AIRs in this paper are often presented as *capacity lower bounds*.

## REFERENCES

- [1] ITU-T, “Forward error correction for submarine systems,” recommendation G.975, 2000, itu.int/rec/T-REC-G.975.
- [2] —, “Forward error correction for high bit-rate DWDM submarine systems,” recommendation G.975.1, 2004, itu.int/rec/T-REC-G.975.1.
- [3] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis and Design*. Chichester, UK: Wiley, 2015.
- [4] L. Schmalen, A. Alvarado, and R. Rios-Müller, “Performance prediction of nonbinary forward error correction in optical transmission experiments,” *J. Lightw. Technol.*, vol. 35, no. 4, pp. 1015–1027, Feb. 2017.
- [5] A. Leven, F. Vacondio, L. Schmalen, S. ten Brink, and W. Idler, “Estimation of soft FEC performance in optical transmission experiments,” *IEEE Photon. Technol. Lett.*, vol. 23, no. 20, pp. 1547–1549, Oct. 2011.
- [6] 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.
- [7] J. Cho and P. J. Winzer, “Probabilistic constellation shaping for optical fiber communications,” *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.
- [8] T. Yoshida, A. Alvarado, M. Karlsson, and E. Agrell, “Post-FEC BER benchmarking for bit-interleaved coded modulation with probabilistic shaping,” *J. Lightw. Technol.*, vol. 38, no. 16, pp. 4292–4306, Aug. 2020.
- [9] E. Agrell *et al.*, “Software supplements,” codes.se/software.
- [10] E. Agrell and M. Secondini, “Information-theoretic tools for optical communications engineers,” in *Proc. IEEE Photon. Conf. (IPC)*, Reston, VA, Sept.–Oct. 2018, pp. 99–103.
- [11] G. Liga, A. Alvarado, E. Agrell, and P. Bayvel, “Information rates of next-generation long-haul optical fiber systems using coded modulation,” *J. Lightw. Technol.*, vol. 35, no. 1, pp. 113–123, Jan. 2017.
- [12] M. Secondini, E. Agrell, E. Forestieri, D. Marsella, and M. R. Camara, “Nonlinearity mitigation in WDM systems: Models, strategies, and achievable rates,” *J. Lightw. Technol.*, vol. 37, no. 10, pp. 2270–2283, May 2019.
- [13] A. Sheikh, A. Graell i Amat, and G. Liva, “Achievable information rates for coded modulation with hard decision decoding for coherent fiber-optic systems,” *J. Lightw. Technol.*, vol. 35, no. 23, pp. 5069–5078, Dec. 2017.
- [14] A. Graell i Amat and L. Schmalen, “Forward error correction for optical transponders,” in *Springer Handbook of Optical Networks*, B. Mukherjee *et al.*, Ed. Cham, Switzerland: Springer Nature, 2020, ch. 7.
- [15] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, “Achievable information rates for fiber optics: Applications and computations,” *J. Lightw. Technol.*, vol. 36, no. 2, pp. 424–439, Jan. 2018.
- [16] S. ten Brink, G. Kramer, and A. Ashikhmin, “Design of low-density parity-check codes for modulation and detection,” *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.
- [17] T. Yoshida, M. Karlsson, and E. Agrell, “Performance metrics for systems with soft-decision FEC and probabilistic shaping,” *IEEE Photon. Technol. Lett.*, vol. 29, no. 23, pp. 2111–2114, Dec. 2017.
- [18] G. Böcherer, P. Schulte, and F. Steiner, “Probabilistic shaping and forward error correction for fiber-optic communication systems,” *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, Jan. 2019.
- [19] A. Sheikh, A. Graell i Amat, G. Liva, and F. Steiner, “Probabilistic amplitude shaping with hard decision decoding and staircase codes,” *J. Lightw. Technol.*, vol. 36, no. 9, pp. 1689–1697, May 2018.
- [20] M. Ivanov, F. Brännström, A. Alvarado, and E. Agrell, “On the exact BER of bit-wise demodulators for one-dimensional constellations,” *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1450–1459, Apr. 2013.
- [21] T. Koike-Akino, D. S. Millar, K. Kojima, and K. Parsons, “Phase noise-robust LLR calculation with linear/bilinear transform for LDPC-coded coherent communications,” in *Proc. Conference on Lasers and Electro-Optics (CLEO)*, San Jose, CA, May 2015.
- [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.

<sup>2</sup>Specifically, we set  $\log q(\mathbf{y}, \mathbf{s})$  equal to the right-hand side of [21, Eq. (3)], where  $R = (3y_c - s_c)/2$ ,  $S = (y_c + s_c)/2$ , and  $y_c$  and  $s_c$  are the complex symbols whose real and imaginary parts are the two components of  $\mathbf{y}$  and  $\mathbf{s}$ .