On achievable rates for long-haul fiber-optic communications

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

On achievable rates for long-haul
fiber-optic communications

Tobias Fehenberger,[1,*] **Alex Alvarado,**[2] **Polina Bayvel,**[2] **and**
Norbert Hanik[1]

[1]*Institute for Communications Engineering, Technische Universität München, 80333 Munich,*
Germany
[2]*Optical Networks Group, University College London (UCL), London, WC1E 7JE, UK*
[*]*tobias.fehenberger@tum.de*

Abstract: Lower bounds on mutual information (MI) of long-haul optical fiber systems for hard-decision and soft-decision decoding are studied. Ready-to-use expressions to calculate the MI are presented. Extensive numerical simulations are used to quantify how changes in the optical transmitter, receiver, and channel affect the achievable transmission rates of the system. Special emphasis is put to the use of different quadrature amplitude modulation formats, channel spacings, digital back-propagation schemes and probabilistic shaping. The advantages of using MI over the prevailing Q-factor as a figure of merit of coded optical systems are also highlighted.

OCIS codes: (060.4080) Modulation; (060.2330) Fiber optics communications.

References and links

1. R.-J. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks," Proceedings of the IEEE **100**(5), 1035–1055 (2012).
2. ITU, *Rec. G.975.1: Forward error correction for high bit-rate DWDM submarine systems* (2004).
3. J. G. Proakis and M. Salehi, *Digital Communications* (McGraw-Hill, 2008), 5th ed.
4. 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. **23**(20), 1547–1549 (2011).
5. M. Karlsson and E. Agrell, "Four-dimensional optimized constellations for coherent optical transmission systems," Proc. European Conference on Optical Communication (ECOC), Paper We.8.C.3 (2010).
6. P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications." Nature **411**, 1027–1030 (2001).
7. I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," J. Lightw. Technol. **23**(11), 3755–3763 (2005).
8. R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," J. Lightw. Technol. **28**(4), 662–701 (2010).
9. B. P. Smith and F. R. Kschischang, "A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications," J. Lightw. Technol. **30**(13), 2047–2053 (2012).
10. 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. **26**(23), 2407–2410 (2014).
11. 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. **31**(23), 3839–3852 (2013).
12. T. Fehenberger and N. Hanik, "Digital back-propagation of a superchannel: Achievable rates and adaption of the GN model," Proc. European Conference on Optical Communication (ECOC), Paper We.3.3.6 (2014).
13. T. Fehenberger, G. Böcherer, A. Alvarado, and N. Hanik, "LDPC coded modulation with probabilistic shaping for optical fiber systems," Proc. Optical Fiber Conference (OFC), Paper Th2A.23 (2015).
14. D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," IEEE Trans. Inf. Theory **52**(8), 3498–3508 (2006).

15. P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," J. Lightw. Technol. **32**(4), 694–721 (2014).

16. U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," IEEE Trans. Inf. Theory **45**(8), 1361–1391 (1999).

17. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, 2006), 2nd ed.

18. A. Alvarado, D. J. Ives, S. J. Savory, and P. Bayvel, "On optimal modulation and FEC overhead for future optical networks," Proc. Optical Fiber Conference (OFC), Paper Th3E.1 (2015).

19. G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, "On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission," Opt. Express **22**(24), 30053–30062 (2014).

1. Introduction

The demand for increased data rates in optical long-haul communications has been growing for several years [1]. For currently deployed fibers, high-order modulation formats are one of the most popular alternatives to increase data rates. The decreased sensitivity of these formats conflicts with the requirement that the bit-error rate (BER) after decoding must be in the order of 10^{-15}. One potential solution to meet this requirement is to use stronger forward error correction (FEC). These advanced FEC schemes typically operate with soft-decision (SD) decoders, i.e., the decoders are fed with soft-information (reliabilities) on the code bits.

Codes with hard-decision (HD) decoding were the de-facto standard in optical long-haul systems [2] as they performed well with limited decoding complexity. Assuming ideal interleaving, the achievable rates of HD-FEC decoders can be fully determined by the pre-FEC BER (or equivalently the Q-factor) [3, Sec. 6.5]. On the other hand, such a relationship between pre-FEC BER and achievable rate does not exist for coded systems with SD decoding. This follows from the fact that SD-FEC decoders use information on the reliability of the bits rather than on HDs. Mutual information (MI) is an achievable transmission rate for SD-FEC decoders, and thus, a natural figure of merit to consider.

In optical communications, MI has been used as a predictor of post-FEC BER that is more reliable than pre-FEC BER [4] and as a reference for the performance of capacity-achieving codes [5]. It has also been used to analytically state lower-bound estimates on capacity [6]. In this paper, we are interested in obtaining a lower bound on MI from the actual output of an optical fiber channel. The objective is to use the bound to quantify achievable rates and compare them for varying system parameters using a Monte Carlo approach. MI estimation from samples of the optical channel has been investigated before. In [7], an estimate of MI for fiber optics is found via computationally extensive simulations. In [8], ring modulations for the optical channel are studied to estimate MI, from which a lower bound on capacity is obtained. [9] and [10] both study probabilistic shaping schemes with MI as figure of merit and compare their respective results with [8]. In [11, Sec. IV-B], equations for lower bounds on MI are presented for continuous channel inputs and quadrature phase-shift keying (QPSK), which are then used to verify a channel model. We follow a similar approach as in [11] to obtain a lower bound, yet provide a simple and general expression for arbitrary discrete modulation formats and use it to analyze optical fiber systems. In this paper, we extend our previous work [12] by formally studying a bound rather than an estimate, and by considering a broader range of system parameters. The bound is obtained with circularly symmetric Gaussian noise statistics, which have been used for this purpose before, e.g., in [11]. We also consider achievable rates for nonuniform input distributions, which is an extension of [13]. In this work, we neglect any memory in the channel, which gives a lower bound on the MI of the true channel with memory [8, Sec. III-F]. We argue that this is a valid assumption as most practical receivers neglect memory by not operating on sequences of symbols but making decisions on a symbol-by-symbol basis.

In this letter, we study achievable rates of a long-haul optical fiber systems. The main con-
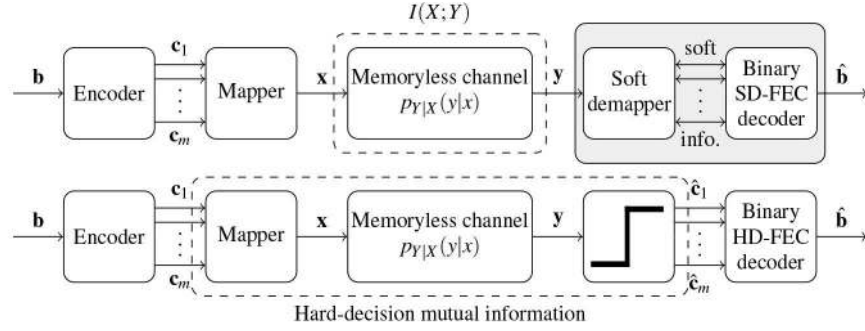
Fig. 1. Block diagram of a coded communication system with SD (top) and HD (bottom) demapping and binary decoding.

tribution is to present ready-to-use expressions for calculating a bound on MI that is based on the input and output symbols and obtained by assuming circularly symmetric Gaussian noise statistics. This allows us to quantify changes in data rate and spectral efficiency when system parameters or the digital signal processing (DSP) of the considered system are varied. We also examine MI for HD decoding and different modulation formats, channel spacings and nonuniform input.

2. Mutual information analysis

2.1. System model

The coded communication system we consider in this paper is shown in Fig. 1. A binary encoder at the transmitter adds redundancy to the information bits $\mathbf{b}$. The m streams of coded bits $\mathbf{c}_i$ are mapped onto symbols that are drawn from a discrete constellation with cardinality $|\mathscr{X}| \triangleq M = 2^m$ and probability mass function (PMF) $P_X(x)$. The sequence of symbols $\mathbf{x}$ is transmitted over a memoryless channel with distribution $p_{Y|X}(y|x)$ and the continuous channel output $\mathbf{y}$ is input into a demapper. The top part of Fig. 1 is an SD decoding system where the soft demapper uses more than two quantization levels. This soft information on the code bits is passed to the binary SD-FEC decoder. The bottom part of Fig. 1 shows the system with an HD demapper and a binary HD-FEC decoder.

2.2. Mutual information as achievable rate

Let Y be the continuous complex output of a memoryless channel with discrete complex input X. The MI $I(X;Y)$ represents the amount of information in bits per channel use (or equivalently, bits per symbol) about X that is contained in Y. The MI is defined as

$$I(X;Y) = \sum_{x \in \mathscr{X}} P_X(x) \int_{\mathbb{C}} p_{Y|X}(y|x) \log_2 \frac{p_{Y|X}(y|x)}{p_Y(y)} dy, \tag{1}$$

which is at most m bits per symbol (bit/sym). $\mathbb{C}$ denotes the set of complex numbers. The operational meaning of MI is an achievable transmission rate; for a fixed PMF $P_X(x)$ and a fixed memoryless channel, it is the largest achievable rate. This means that when we transmit below this rate, coding schemes exist that allow the post-FEC BER to be made arbitrarily small. The converse is also true: An arbitrarily small post-FEC BER cannot be achieved when we transmit at a rate larger than the MI. Higher transmission rates require changing the PMF $P_X(x)$ or the channel. Both of these options are analyzed in Sec. 3 for a multi-span wavelength-division multiplexing (WDM) system.

2.3. Lower bounds on mutual information

In order to calculate Eq. (1), the channel transition probability $p_{Y|X}(y|x)$ must be known. Since no analytical expression exists for an optical fiber channel, we need to bound Eq. (1). We obtain a lower bound on Eq. (1) by using the mismatched decoding approach [14, Sec. VI] and consider an auxiliary channel with transition probability $q_{Y|X}(y|x)$ instead of the memoryless $p_{Y|X}(y|x)$. The output distribution of the auxiliary channel is $q_Y(y) = \sum_{x \in \mathscr{X}} q_{Y|X}(y|x)P_X(x)$. We bound $I(X;Y)$ in Eq. (1) as

$$I(X;Y) \geq \mathrm{R_{SD}} \triangleq \sum_{x \in \mathscr{X}} P_X(x) \int_{\mathbb{C}} p_{Y|X}(y|x) \log_2 \frac{q_{Y|X}(y|x)}{q_Y(y)} \mathrm{d}y. \qquad (2)$$

$\mathrm{R_{SD}}$ is an achievable rate for a receiver with SD-FEC decoding as shown in Fig 1 (top). In this work, the transition probability $q_{Y|X}(y|x)$ is taken to be Gaussian with noise variance σ^2:

$$q_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right). \qquad (3)$$

We justify the assumption of white Gaussian noise statistics by the good agreement of fiber simulations shown in [12] and the Gaussian noise model [15]. We approximate $\mathrm{R_{SD}}$ in Eq. (2) using a Monte Carlo approach. Suppose we receive N distorted complex symbols $y_n, n = 1, \ldots, N$. We first estimate the noise variance σ^2 from all received symbols. The conditional probability that a specific x was sent when we observe y_n follows from Eq. (3) and Bayes' theorem:

$$q_{X|Y}(x|y_n) = \frac{\exp\left(-\frac{(y_n-x)^2}{2\sigma^2}\right)P_X(x)}{\sum_{x' \in \mathscr{X}} \exp\left(-\frac{(y_n-x')^2}{2\sigma^2}\right)P_X(x')}. \qquad (4)$$

We calculate $q_{X|Y}(x|y_n)$ for all $M \cdot N$ combinations of potential x and observed y_n. A ready-to-use expression for the lower bound on MI is then

$$\mathrm{R_{SD}} \approx -\sum_{x \in \mathscr{X}} P_X(x) \log_2 P_X(x) + \frac{1}{N} \sum_{n=1}^{N} \sum_{x \in \mathscr{X}} q_{X|Y}(x|y_n) \log_2 q_{X|Y}(x|y_n). \qquad (5)$$

For uniformly distributed input, $P_X(x) = \frac{1}{M}$ and Eq. (5) simplifies to

$$\mathrm{R_{SD}} \approx m + \frac{1}{N} \sum_{n=1}^{N} \sum_{x \in \mathscr{X}} \frac{\exp\left(-\frac{(y_n-x)^2}{2\sigma^2}\right)}{\sum_{x' \in \mathscr{X}} \exp\left(-\frac{(y_n-x')^2}{2\sigma^2}\right)} \log_2 \frac{\exp\left(-\frac{(y_n-x)^2}{2\sigma^2}\right)}{\sum_{x' \in \mathscr{X}} \exp\left(-\frac{(y_n-x')^2}{2\sigma^2}\right)}. \qquad (6)$$

The estimation accuracy of Eqs. (5) and (6) increases with N.

Achievable rates for the binary HD-FEC shown in Fig. 1 (bottom) are calculated by considering two different FEC schemes. The first setup consists of m parallel binary encoder-decoder pairs, which means that there is a component code for each of the m binary sub-channels. We consider a multistage decoder setup with no information exchange between the component decoders [16, Sec. 6]. Let $C_i \ldots C_m$ be binary random variables that are input into the mapper and $\hat{C}_i \ldots \hat{C}_m$ the corresponding binary inputs into the binary decoder. An achievable rate for this design is denoted by $\mathrm{R_{HD}^m}$ and defined as

$$\mathrm{R_{HD}^m} \triangleq \sum_{i=1}^{m} I(C_i; \hat{C}_i) = \sum_{i=1}^{m} (1 - H_{\mathrm{b}}(p_i)), \qquad (7)$$

where $H_b(p_i) = -p_i \log_2 p_i - (1-p_i)\log_2(1-p_i)$ is the binary entropy function and p_i is the BER of the ith binary symmetric channel with input C_i and output $\hat{C}_i$. The second case we consider consists of one binary encoder-decoder pair that operates jointly on all m sub-channels and thus "sees" the average bit error rate $\bar{p} = \sum_{i=1}^{m} p_i/m$. An achievable rate for this FEC scheme is denoted by R_{HD}^1 and defined as

$$R_{\text{HD}}^1 \triangleq m \cdot (1 - H_b(\bar{p})) \leq R_{\text{HD}}^m, \tag{8}$$

where the inequality follows from Jensen's inequality [17, Sec. 2.6].

3. Mutual information analysis for long-haul optical fiber communications

3.1. Simulation Setup

The dual-polarization multi-span WDM simulation setup we consider here is shown in Fig. 2. We generate 2^{16} Gray-labeled QAM symbols per polarization from a random bit sequence. A root-raised cosine (RRC) filter with 5% roll-off is used for pulse-shaping. The pulses are ideally converted into the analog and optical domain. The symbol rate is 28 GBaud and the signal bandwidth is 29.4 GHz. The number of simulated WDM channels N_{ch} varies with the channel spacing B_{ch} such that the total signal bandwidth is kept constant at 450 GHz. The above steps are repeated to generate the y-polarization.

The simulated fiber link consists of N_s spans of length 100 km, each followed by an Erbium-doped fiber amplifier (EDFA) with a noise figure of 4 dB. The fiber is single-mode fiber (SMF) with α=0.2 dB/km, γ=1.3 (W·km)$^{-1}$ and D=17 ps/nm/km. Signal propagation is simulated using the split-step Fourier method with 32 samples per symbol. The step size is 100 m for the linear regime and decreased to 10 m for the nonlinear (high input power) regime.

The incoming optical signal is filtered with an ideal optical band-pass filter and converted ideally into the digital domain. Either ideal electronic dispersion compensation (EDC) or ideal digital back-propagation (DBP) with the same step size and samples per symbol as for the forward propagation are applied. We either use single-channel (SC) DBP or multi-channel (MC) DBP of the full field. Neither equalization nor carrier phase recovery are required as polarization mode dispersion and laser phase noise are not present. After matched filtering and down-sampling, R_{SD} is calculated from the received symbols of the center channel as outlined in Sec. 2.3 and averaged over both polarizations. The dashed box in Fig. 2 depicts the channel for which R_{SD} is calculated. The channel contains not only the fiber itself but also most of the transmitter and receiver DSP. R_{HD}^m and R_{HD}^1 are calculated from the pre-FEC BERs of the center WDM channel after hard-decision demapping.
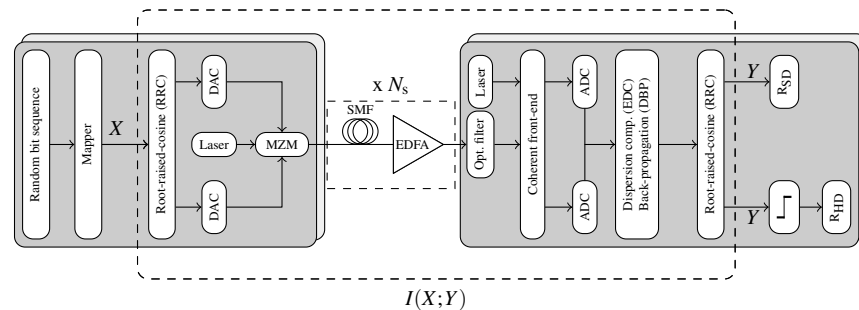

Fig. 2. Block diagram of the simulated optical system. The dashed box includes all components and subsystems that influence $I(X;Y)$.
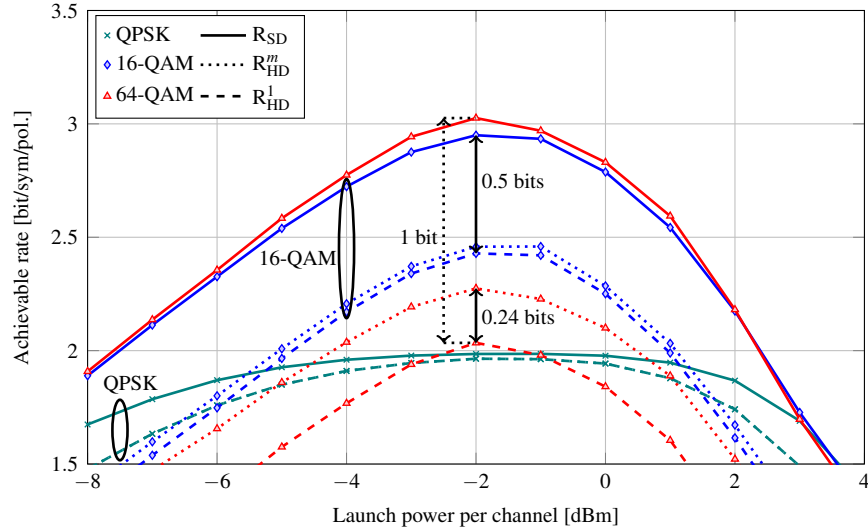
Fig. 3. Achievable rates for soft- and hard-decision decoding: R_{SD} (solid), R_{HD}^m (dotted), R_{HD}^1 (dashed) and different modulation formats: QPSK (green, x), 16-QAM (blue, diamond) and 64-QAM (red, triangle). EDC only, 6000 km SMF and N_{ch}=15 WDM channels.

3.2. Modulation format

In Fig. 3, the achievable rates of three modulation formats vs. launch power are compared for N_{ch}=15 WDM channels, EDC, B_{ch}=30 GHz and 6000 km link length (N_s=60). When considering the SD rate R_{SD} (solid lines), QPSK reaches a plateau that is very close to its maximum of 2 bit/sym. This suggests that QPSK is not the best choice for the input X and we should use a higher-order QAM. The gain in MI of 64-QAM over 16-QAM at their respective optimum power is only about 0.1 bit/sym. Furthermore, 64-QAM imposes stronger requirements on digital-to-analog converters and the optical signal processing algorithms than 16-QAM and is expected to suffer from a larger implementation penalty. We therefore conclude that 16-QAM is the optimal square QAM for the setup under consideration.

The operational meaning of the maximum 16-QAM MI of 2.95 bit/sym is as follows. For the given setup and a receiver operating under the assumption of Gaussian and independent received symbols, at most 2.95 bits of the 4 bits of each 16-QAM symbol are available for transmitting information. An ideal FEC operating at this rate must have a coding rate of at most 2.95/4=0.7375, which corresponds to a coding overhead of 35.6%. A higher transmission rate (or less overhead) can under no circumstances give a post-FEC BER in the order of 10^{-15}.

Figure 3 also depicts the results for HD decoding. Let us first compare R_{HD}^m (dotted lines) and R_{HD}^1 (dashed lines). For QPSK, $R_{HD}^m = R_{HD}^1$ as two orthogonal binary channels are effectively considered. The rate improvement by having m encoder-decoder pairs is less than 0.05 bit/sym for 16-QAM and about 0.24 bit/sym for 64-QAM. We observe that the difference between R_{HD}^m and R_{HD}^1 grows with modulation order because averaging over more parallel subchannels means that more information is lost. In the following, we consider only R_{HD}^1 as it is an achievable rate for the practically relevant case of one FEC encoder-decoder pair [9].

Comparing R_{SD} and R_{HD}^1 (dashed lines), Fig. 3 illustrates that the gap between R_{SD} and R_{HD}^1 grows with modulation order. For QPSK, the gap is 0.05 bit/sym, it increases to 0.5 bit/sym for 16-QAM, and is up to 1 bit per symbol for 64-QAM. This is because the potential gain of SD over HD depends on the considered coding rate; it vanishes for high coding rates [3, Sec. 6.8]. The conclusion is that the closer a system is operated to its maximum transmission rate of m,
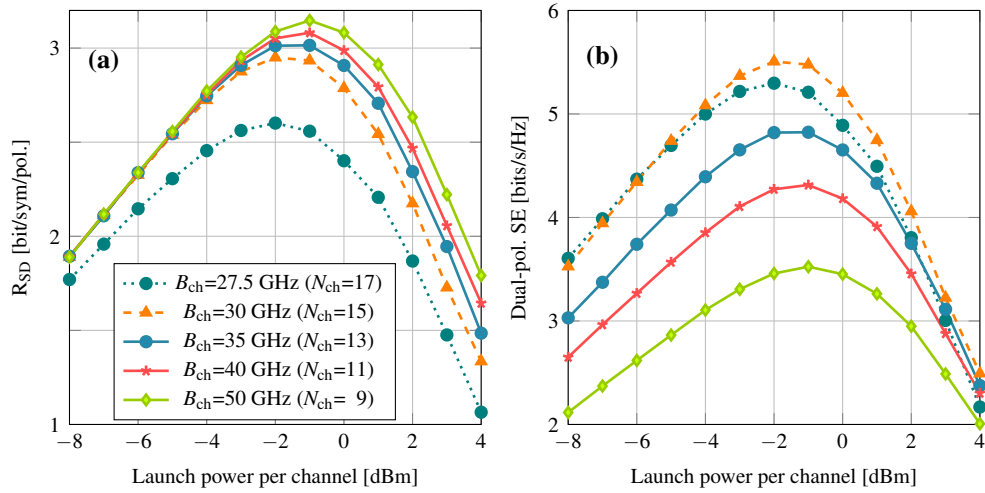
Fig. 4. Achievable rates R_{SD} (a) and dual-polarization SE (b) for 16-QAM and varying WDM channel spacings B_{ch}. The total bandwidth over all WDM channels is constant at 450 GHz and the signal bandwidth per WDM channel is 29.4 GHz.

the less beneficial SD decoding becomes. The importance of MI becomes evident in this case as it allows us to draw the conclusion that, when we only consider the achievable rates and disregard practical aspects of modulation formats, R_{SD} suggests to use 64-QAM, while R_{HD}^1 tells us that 16-QAM is the best choice. Note that R_{HD}^1 will eventually be larger for 64-QAM than for 16-QAM when 16-QAM is operated close to 4 bit/sym, see, e.g., [18, Fig. 2]. The Q-factor does not allow this kind of analysis.

3.3. WDM spacing

In Fig. 4 (a), R_{SD} is evaluated for 16-QAM, EDC, 6000 km and N_{ch} WDM channels spaced at B_{ch} from 27.5 GHz to 50 GHz. For 30 GHz (dashed curve), which is close to the Nyquist rate of 29.4 GHz, 2.95 bit/sym are achievable at the optimal power. Increasing the channel spacing from 30 GHz to 50 GHz (solid curves) results in a larger rate of up to 3.15 bit/sym at the optimum power because the phase mismatch between co-propagating signals increases and the impact of four-wave mixing decreases. In the linear regime, the MI does not change for spacings from 30 GHz to 50 GHz. If the channel spacing is reduced to a sub-Nyquist 27.5 GHz (dotted line), the spectral overlap between adjacent WDM channels leads to additional noise that decreases R_{SD} also in the linear regime.

In the context of constant signal bandwidth and varying WDM spacings, spectral efficiency (SE) must be considered as it directly relates to the net data rate of the system by taking into account the bandwidth usage. We calculate SE from the per-polarization rates R_{SD} of Fig. 4 (a) as $2 \cdot R_{SD} \cdot 28$ GBaud/B_{ch}. We emphasize that we consider only the center channel, i.e., the one that experiences the strongest nonlinear interference. This is a worst-case scenario and higher rates might be achievable in other channels. Results are shown in Fig. 4 (b). A maximum SE of 5.51 bit/s/Hz is obtained for the quasi-Nyquist spacing of 30 GHz. In this case, almost the entire spectrum of a WDM channel is used, which increases the SE more than the smaller MI decreases it. We conclude that there is no need to have large spectral guard bands to obtain the maximum SE of the center channel.
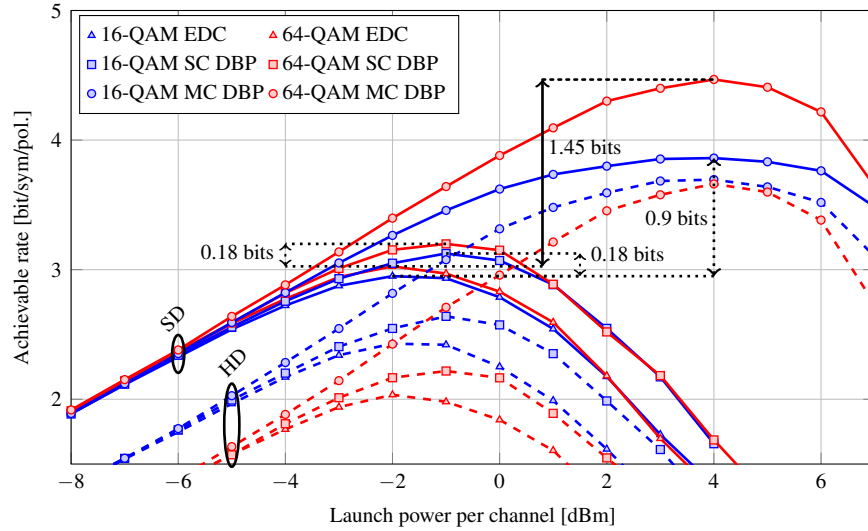
Fig. 5. Achievable rates for soft- and hard-decision decoding: R_{SD} (solid) and R_{HD}^1 (dashed) and different modulation formats: 16-QAM (blue) and 64-QAM (red). The simulation setup is identical to the one in Fig. 3.

3.4. EDC, single-channel DBP and multi-channel DBP

In Fig. 5, EDC, SC DBP and MC DBP are compared for the same parameter set as for Fig. 3. By replacing EDC with SC DBP, R_{SD} is increased by 0.18 bit/sym for both 16-QAM and 64-QAM. Similar gains are found for R_{HD}^1. 16-QAM with MC DBP and SD decoding gives a gain of 0.9 bit/sym over EDC. The plateau around 4 dBm suggests that larger gains are possible with 64-QAM. Our simulations confirm that an increase of 1.45 bit/sym is possible for 64-QAM when MC DBP is used instead of EDC. Note that the slopes of the curves in Fig. 5 in the highly nonlinear regime are less steep than previously reported in [12]. This is explained by the fact that here we use a step size of 10 m while in [12] 100 m was used. A comparison of ideal SC DBP and MC DBP similar to Fig. 5 was performed in [19] using the Q-factor as figure of merit. The results in [19] are similar to the ones in Fig. 5. For SD decoding, however, only a qualitative comparison of different DBP schemes is possible with the Q-factor, and a relation to data rates gain cannot be made. For MC DBP, the difference between R_{SD} and R_{HD}^1 is 0.17 bit/sym for 16-QAM and 0.8 bit/sym for 64-QAM. A practical interpretation of this is that for 16-QAM with MC DBP and the chosen system parameters, the extra complexity of SD-FEC might be spared because the potential improvement of SD-FEC over HD-FEC is small when R_{HD}^1 is close to m.

3.5. Probabilistic shaping for 64-QAM

In this section, we choose an input X that is not uniformly distributed over $\mathcal{X}$, but instead, we probabilistically shape the constellation using a heuristic scheme. As explained in [13], the input PMF is chosen to be the Maxwell-Boltzmann distribution that maximizes MI for an additive white Gaussian noise channel under a power constraint. The rate R_{SD} in this case is shown in Fig. 6 as a function of transmission distance. Each point in this plot is obtained for the optimum launch power and N_{ch}=15 WDM channels. We see in Fig. 6 that for 64-QAM and EDC only, transmission over additional 200 km is possible without sacrificing data rate by shaping the input. For distances above 1000 km, the shaping gain over uniform input lies
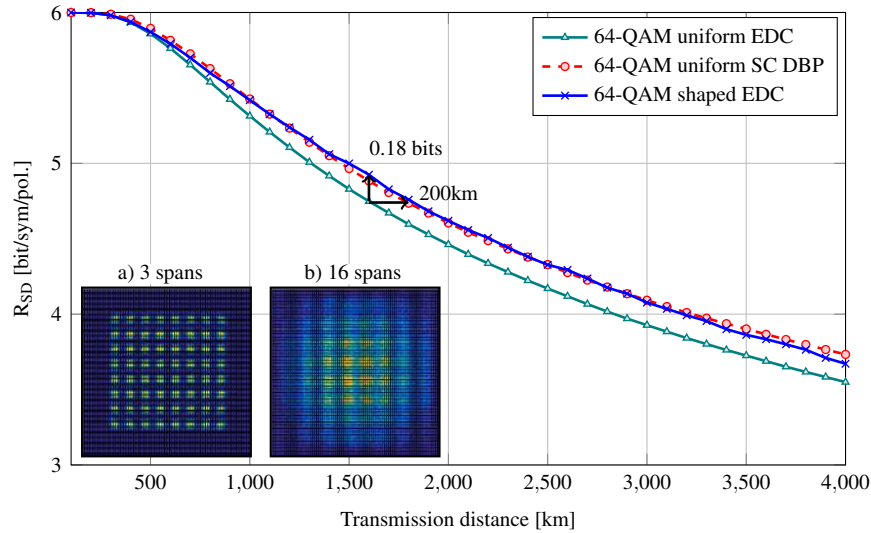
Fig. 6. Shaped 64-QAM outperforms uniform 64-QAM (both with EDC) in rate and distance and gives similar gains as uniform input with SC DBP. The insets a) and b) show the shaped received constellation after 3 and 16 spans, respectively.

between 0.1 bit/sym and 0.2 bit/sym, which is comparable to the gain by SC DBP at its optimal launch power. The fact that the employed heuristic shaping scheme performs as good as the computationally extensive ideal SC DBP shows the great potential of probabilistic shaping.

4. Conclusion

We have studied MI as a figure of merit to gain insights into coded fiber-optic systems. Unlike the Q-factor, MI represents an achievable rate, and thus, it can be directly connected to the spectral efficiency of the system. Using MI as the figure of merit enables a more meaningful analysis as it allows to quantify changes in the spectral efficiency of the system. Although some conclusions drawn in Sec. 3 apply specifically to the presented configuration, the application of the MI as a tool for system design is not limited to this configuration, but can be extended to any device and algorithm of the optical channel, including transmitter and receiver DSP.

We have lower-bounded the true MI of the optical channel by making two assumptions. We have first neglected any memory and have then used the mismatched decoder approach with white Gaussian noise statistics. An interesting question is how much is to be gained by dropping these two simplifying assumptions. In particular, we believe that tighter lower bounds on MI can be found in systems for which the Gaussian noise assumption does not hold, such as single span systems, or at high launch powers. These questions are left for further investigation.

Acknowledgment