Information-Theoretic Tools for Optical Communications Engineers [Invited]

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

Agrell, E., Secondini, M. (2018). Information-Theoretic Tools for Optical Communications
Engineers [Invited]. 31st Annual Conference of the IEEE Photonics Society, IPC 2018.
http://dx.doi.org/10.1109/IPCon.2018.8527126

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

(article starts on next page)

Information-Theoretic Tools for Optical Communications Engineers

(Invited Tutorial)

Erik Agrell
Department of Electrical Engineering
Chalmers University of Technology
SE-41296 Gothenburg, Sweden
agrell@chalmers.se

Marco Secondini
TeCIP Institute
Scuola Superiore Sant'Anna
I-56124 Pisa, Italy
marco.secondini@sssup.it

Abstract—**Fundamental information-theoretic concepts are explained for nonspecialists, with emphasis on their practical usage. The notions of a "FEC threshold" and a "nonlinear Shannon limit" are critically reviewed, highlighting their limitations and possible alternatives.**

Index Terms—**Achievable information rate, bit error rate, block error rate, channel capacity, FEC limit, FEC threshold, fiber-optic communications, forward error-correction coding, generalized mutual information, nonlinear Kerr effect.**

I. INTRODUCTION

By adding a controlled amount of redundancy to a transmitted digital signal, and checking this redundancy in the receiver, transmission errors can be corrected. This is the idea behind *forward error correction* (FEC) coding, which is an integral part of almost any digital communication system, including fiber-optic systems. A fundamental question is how much FEC overhead (OH) is required to guarantee a certain performance. Claude E. Shannon found in 1948, somewhat counter-intuitively, that error-free performance (in the sense of arbitrarily small bit or block error rate) can be achieved over *any* noisy channel, provided that data is arranged in long enough blocks and that ideal FEC with sufficient OH is applied to each block [1]. In the limit of infinite block length, the achievable bit error rates (BER) and block error rates (BLER) are shown in Fig. 1 as functions of the information rate (which depends on the OH). The threshold rate, below which error-free performance is achievable, is defined as the *channel capacity C*. Interestingly, the regions have exactly the same shape for any channel [2, p. 195], although the channel capacity may differ. These results and their far-reaching extensions, which constitute the field of *information theory,* apply also to nonlinear media such as optical fibers, although they were conceived much later.

II. THE FIBER NONLINEARITY AND ITS IMPLICATIONS

The reference model for the propagation of light in optical fibers is the nonlinear Schrödinger equation (NLSE) [3, p. 50]
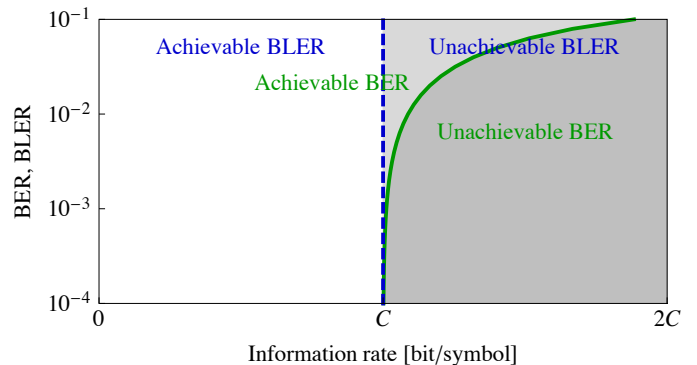
Fig. 1. Achievable bit and block error rates as a function of information rate, at asymptotically large block length. Any BER to the left of the solid green curve can be achieved for any communication channel, and so can any BLER to the left of the dashed blue curve. The capacity C is the highest rate for which an arbitrarily small BER or BLER is achievable.

(or its two-dimensional version, the Manakov equation, in the polarization-multiplexed case [4]). Besides attenuation and dispersion, the NLSE includes Kerr nonlinearity, a peculiar characteristic of the optical fiber channel which determines a rotation of the phase of the optical signal proportional to its instantaneous intensity. According to the split-step method [3, Ch. 2], the NLSE channel can be practically represented by dividing the fiber link into many small steps, over which the effects of dispersion and Kerr nonlinearity are split, as shown in the green block in Fig. 2. Noise from optical amplifiers is also introduced at every step (for distributed amplification).

With no or negligible Kerr nonlinearity (e.g., at low power), the classical theory of additive white Gaussian noise (AWGN) channels could be essentially applied, and capacity would depend on the signal-to-noise ratio (SNR) through the well known $\log(1 + \text{SNR})$ formula [1]. With Kerr nonlinearity, on the other hand, the problem becomes very complicated: signal and noise interact nonlinearly and non-instantaneously during propagation, giving rise to non-Gaussian channel statistics, long temporal correlations (memory), and spectral broadening. The situation gets even more complicated in wavelength-division-multiplexed (WDM) systems, in which nonlinear in-
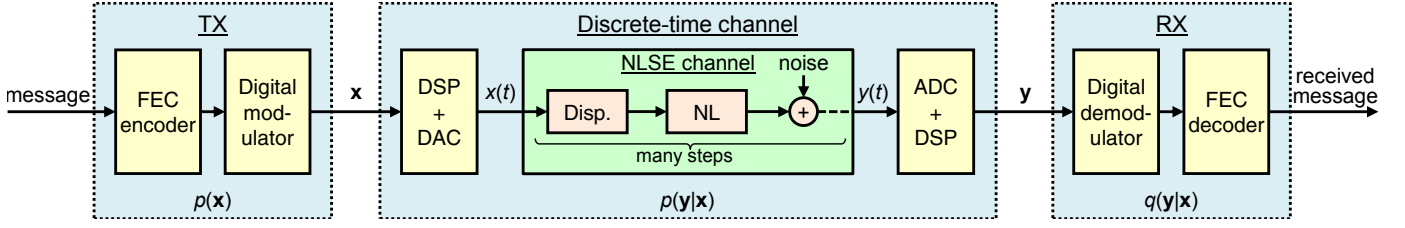
Fig. 2. Schematic representation of an optical fiber system, showing the split-step model of the NLSE channel, including dispersion, nonlinearity (NL), and noise, and its discrete-time version seen by the TX and RX.

teractions take place also between WDM signals (not represented in Fig. 2 for simplicity). For these reasons, channel statistics cannot be analytically evaluated nor reliably estimated numerically, so that the capacity of the optical fiber at high power is essentially unknown.

An optical fiber communication system can be conveniently represented by the model depicted in Fig. 2. A sequence of information bits (the message) is processed by a FEC encoder, which adds some redundant bits needed for error correction, and mapped to a sequence of symbols $\mathbf{x} = (x_1, \ldots, x_N)$ by a digital modulator; symbols $\mathbf{x}$ are then encoded in a waveform $x(t)$ by digital signal processing (DSP) and a digital-to-analog converter (DAC) (e.g., a linear modulator); $x(t)$ propagates through the NLSE channel; the corresponding output waveform $y(t)$ is filtered and sampled by an analog-to-digital converter (ADC) and DSP block to obtain the output samples $\mathbf{y} = (y_1, \ldots, y_N)$; the output samples are finally demodulated and decoded to detect the transmitted symbols, correct for possible errors, and eventually extract the message.

From an information-theoretic perspective, the system can be first studied at a higher level by considering the three light blue blocks highlighted in Fig. 2: the transmitter (TX), comprising the FEC encoder and digital modulator and characterized by the distribution $p(\mathbf{x})$; the discrete-time channel, comprising the DAC, NLSE channel, and ADC and characterized by the conditional distribution $p(\mathbf{y}|\mathbf{x})$; and the receiver (RX), consisting of the digital demodulator and FEC decoder. Since the actual channel statistics $p(\mathbf{y}|\mathbf{x})$ are unknown, the RX makes decisions based on a possibly mismatched channel law $q(\mathbf{y}|\mathbf{x})$.[1]

A relevant quantity in this case is the *achievable information rate* (AIR) [5], [6]

$$I_q(X;Y) = \lim_{N \to \infty} \frac{1}{N} E \left\{ \log \frac{q(\mathbf{y}|\mathbf{x})}{q(\mathbf{y})} \right\} \quad (1)$$

$$\approx \frac{1}{N} \log \frac{q(\mathbf{y}|\mathbf{x})}{q(\mathbf{y})} \quad \text{[bit/symbol]} \quad (2)$$

where $q(\mathbf{y}) = \int p(\mathbf{x})q(\mathbf{y}|\mathbf{x})d\mathbf{x}$, and *achievable* means that, by using an ideal FEC code (in the sense of the previous section), it is possible to encode $I_q(X;Y)$ information bits per symbol on $\mathbf{x}$, transmit them through the channel of Fig. 2, and reliably

(with arbitrarily low error probability) decode them from $\mathbf{y}$ by using a receiver optimized for $q(\mathbf{y}|\mathbf{x})$. While the expectation in (1) should be computed with respect to the actual joint statistics $p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$, the approximation (2)—based on the asymptotic equipartition property [7, Ch. 3] and accurate for large N—can be readily evaluated both in numerical simulations and experiments without any knowledge of the actual channel statistics. In practice, the knowledge of $p(\mathbf{y}|\mathbf{x})$ is not required either to compute (2) or to design a system that can achieve it. In the context of optical fiber communications, the AIR has been explicitly used as a practical performance metric, both in linear (e.g., [8]–[10]) and nonlinear scenarios (e.g., [11]–[14]).

Based on (1)–(2), the problem of capacity evaluation can be practically subdivided into three parts: the computation of the AIR (2) for given $p(\mathbf{x})$ and $q(\mathbf{y}|\mathbf{x})$; the optimization of (2) over $q(\mathbf{y}|\mathbf{x})$, which can also be regarded as a channel modeling/estimation problem; and the optimization over $p(\mathbf{x})$, which corresponds to an optimization of the modulation format. The maximum over $q(\mathbf{y}|\mathbf{x})$, which always occurs for $q(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\mathbf{x})$ [5], is called the *mutual information*, and the maximum mutual information over $p(\mathbf{x})$, which can only be computed numerically except for very simple channels such as the AWGN, is the *channel capacity* C, which was conceptually defined in the previous section.

Within this framework, any suboptimum choice of $p(\mathbf{x})$ and $q(\mathbf{y}|\mathbf{x})$ leads to an achievable rate and, thus, to a capacity lower bound. For instance, a typical choice is that of optimizing the input distribution and detection metric for an AWGN channel, i.e., drawing independent and identically distributed circularly symmetric Gaussian input symbols $\mathbf{x}$ and assuming that the Kerr nonlinearity can be accounted for in $q(\mathbf{y}|\mathbf{x})$ by simply increasing the variance of the AWGN affecting the output symbols $\mathbf{y}$. This choice, supported by the Gaussian noise model [15] and made (or practically entailed), for instance, in [16], [17], yields an AIR that has the same formal expression as the AWGN channel capacity, $\log(1 + \text{SNR})$, but with an SNR that, rather than increasing unboundedly with power as in true AWGN channels, has a finite maximum at some optimum power, after which it decreases again due to Kerr nonlinearity. The corresponding maximum AIR, often referred to as *nonlinear Shannon limit* [18], has been sometimes overinterpreted as a true capacity limit but, more rigorously, should be considered just as a capacity lower bound achievable by conventional (i.e., optimized for the AWGN channel) systems [19]. In fact, better

---

[1]In this representation, the DAC and ADC, as well as some DSP operations aimed at simplifying the channel model, are formally included in the discrete-time channel, though usually physically located in the TX or RX.
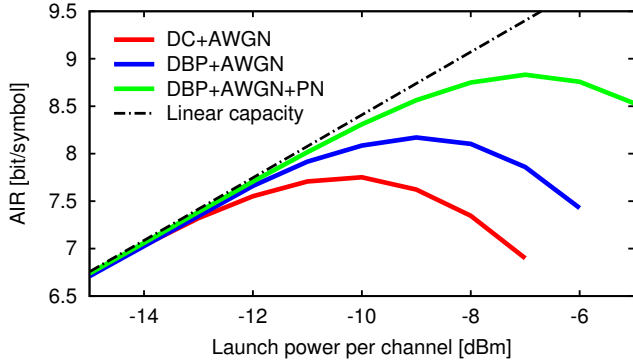
Fig. 3. AIRs for a five-channel WDM system over a 1000 km link with ideal distributed amplification and different detection metrics.

choices of $p(\mathbf{x})$ and $q(\mathbf{y}|\mathbf{x})$ might yield higher AIRs.

As an example, Fig. 3 shows the AIR as a function of the launch power for a five-channel Nyquist-WDM system over 1000 km of standard fiber with ideal distributed amplification (simulated by the split-step method as shown in Fig. 2), a Gaussian input distribution $p(\mathbf{x})$, and different detection metrics $q(\mathbf{y}|\mathbf{x})$: AWGN after dispersion compensation (DC) (neglecting nonlinearity); AWGN after digital backpropagation (DBP) (accounting for deterministic intra-channel nonlinearity); and AWGN plus Wiener phase noise (PN) after DBP (accounting also for PN caused by inter-channel nonlinearity). The second and third metrics, being better matched to the true channel statistics, provide an AIR gain of about 0.4 and 1.1 bit/symbol, respectively. AIR gains based on different choices of $p(\mathbf{x})$ and/or $q(\mathbf{y}|\mathbf{x})$ have been reported for various scenarios (single-channel or WDM, distributed or lumped amplification, single- or dual-polarization, . . .) [14], [19]–[22].

Even the very existence of an AIR maximum at some finite optimum power is related to the suboptimum choice of $p(\mathbf{x})$ and $q(\mathbf{y}|\mathbf{x})$, as channel capacity is a monotonic function of average power [23]. In fact, the unbounded capacity (growing to infinity for infinite power) of many simplified optically-related nonlinear channels has been demonstrated [19], [24], [25].

III. POST-FEC BER PREDICTION

In the development and optimization of optical communication systems, it is common to separate the design and evaluation of the FEC code from the rest of the system. The signal quality is estimated after the optical receiver and DSP but before FEC decoding, and known properties of the code are utilized off-line to predict the performance after decoding. This is because the BER after FEC decoding, the so-called *post-FEC BER*, is typically in the order of 10^{-15} or less, and it is very resource-demanding to estimate such low BER values.

The most common signal quality metric for this purpose is the *pre-FEC BER*. The pre-FEC BER p required to achieve a certain post-FEC BER is called the *FEC threshold* or *FEC limit*. Such FEC thresholds have been computed for a variety of codes assuming a *binary symmetric channel,*

which is a discrete-time, binary channel where bit errors occur independently and with equal probability p for zeros and ones. For this channel, FEC thresholds have been estimated by dedicated FEC-decoding hardware, Monte-Carlo simulation data extrapolated to low post-FEC BERs, or analytic approximations. FEC thresholds of some selected codes that have been recommended for use in optical communication systems are listed in Table I, along with their sources.

In optical experiments and simulations, it is very common to define acceptable (i.e., almost error-free) performance using a FEC threshold. In most literature, however, FEC thresholds are given without a reference, or with an incorrect one. This unfortunate practice has led to numerous questionable claims, in terms of thresholds that cannot be validated and in some cases contradict other published thresholds for the same code. As an illustrative example, the presently most popular FEC threshold, which is $3.8 \cdot 10^{-3}$, was used in at least 44 papers at the Optical Fiber Communication Conference (OFC) 2017 alone, but not once with a correct reference. Three times at OFC 2017, and many more times in other publications, it was incorrectly attributed to [28, App. I.9], where this particular threshold does not appear. Similar error propagation can be observed among references for other FEC thresholds.

When applied in practical communication systems, the validity of FEC thresholds relies on two key assumptions: hard decisions (HD) and independent, uniform pre-FEC bit errors. Otherwise the channel is statistically different from the binary symmetric channel for which the FEC performance was derived, and the performance cannot be guaranteed. We will now briefly discuss these two assumptions in turn.

HD means that the input to the FEC decoder is bits and nothing else. No information about the reliability of these bits is passed from the demodulator. Therefore the FEC thresholds in Table I, which were all derived under this assumption, are sometimes more precisely called *HD-FEC thresholds*. However, p cannot be used for precise post-FEC BER predictions in systems with soft decision (SD) decoding [37]. For this reason, so-called SD-FEC thresholds should be avoided, or at least used with caution. They are approximations only, with unknown accuracy. Two systems, operating on different channels or using different modulation, may have the same pre-FEC BER p but different soft metrics (probabilities that the pre-FEC bits are correctly detected). Such systems will have the same post-FEC BER if HD-FEC is applied but different in case of SD-FEC.

If the pre-FEC BER is unreliable for post-FEC BER prediction in SD systems, which metric should be used instead? The mutual information was proposed as a metric in [38] and the *generalized mutual information* (GMI), defined as the sum of bit-wise mutual informations[2] was proposed in [37]. Of these metrics, the mutual information provides the most accurate prediction in systems with nonbinary coding and symbol-wise hard or soft decisions [40], whereas the GMI

---

[2]This definition of GMI is the dominating one in optical communications, whereas "GMI" often denotes a more general quantity in the information-theoretic literature [39, Ch. 4].

TABLE I
HARD-DECISION FEC THRESHOLDS OF CERTAIN CODES RECOMMENDED FOR OPTICAL COMMUNICATIONS

OH	Code type	Source	post-FEC BER	p
2.7%	Reed–Solomon (RS) (528,514) "KR4"	[26][a]	10^{-15}	$2.18 \cdot 10^{-5}$
5.8%	RS(544,514) "KP4"	[26][a]	10^{-15}	$2.26 \cdot 10^{-4}$
6.7%	RS(255,239)	[27][a]	10^{-15}	$8.27 \cdot 10^{-5}$
6.7%	Bose–Chaudhuri–Hocquenghem (BCH)+BCH	[28, App. I.3]	10^{-15}	$3.15 \cdot 10^{-3}$
6.7%	BCH+RS	[28, App. I.4]	10^{-15}	$2.17 \cdot 10^{-3}$
6.7%	RS(2720,2550)	[28, App. I.8]	10^{-15}	$1.10 \cdot 10^{-3}$
6.7%	Proprietary "P-FEC"	[29]–[32]	$10^{-15} - 10^{-20}$	$3.84 \cdot 10^{-3}$
6.7%	BCH+BCH	[33][b,c]	10^{-15}	$4.42 \cdot 10^{-3}$
7% (6.7%?)	Continuously interleaved BCH	[34]	10^{-15}	$4.52 \cdot 10^{-3}$
~~20%~~	~~Low-density parity-check (LDPC) convolutional~~	~~[35][b]~~	~~10^{-15}~~	~~$2.7 \cdot 10^{-2}$~~
24.5%	Convolutional+RS	[28, App. I.2]	10^{-15}	$5.20 \cdot 10^{-3}$
$6.7 - 24.3\%$	BCH+BCH	[28, App. I.7]**c**	10^{-15}	$1.30 \cdot 10^{-3} - 1.30 \cdot 10^{-2}$
$6.25 - 33.33\%$	Staircase	[36][c]	10^{-15}	$4.70 \cdot 10^{-3} - 2.24 \cdot 10^{-2}$

[a]The pre-FEC BER p was estimated by the accurate approximation in [28, App. I.8.2], assuming bounded-distance decoding.
[b]The pre-FEC BER was calculated from the given Q factor in dB as $p = (1/2)\mathrm{erfc}(10^{Q/20}/\sqrt{2})$.
[c]The pre-FEC BER or Q factor was extrapolated from simulation results at significantly higher post-FEC BERs, and may therefore be inaccurate.
Errata:
  • The LDPC convolutional code in [35] uses soft-decision decoding and should not be included in this table.
  • The BCH+BCH codes in [28, App. I.7] should have a superscript c, since two of the three thresholds come from extrapolated simulations.

gives better prediction in systems with bit-wise soft decisions [37]. Unfortunately, no GMI thresholds for specific FEC codes at low post-FEC BERs such as 10^{-15} have yet been published, but GMI thresholds at higher post-FEC BERs are available, suitable for use in concatenated FEC systems based on an inner SD decoder and and outer HD decoder [37], [40]. The GMI can also be interpreted as an achievable rate with ideal FEC coding.

To address the second key assumption, that bit errors occur independently, a large interleaver is typically applied (or assumed) after the FEC encoder, and a corresponding deinterleaver before the FEC decoder. Otherwise the input bits (and their corresponding soft metrics) to the decoder will be correlated. This correlation may come from several sources, such as residual channel memory, transceiver imperfections, cycle slips in the phase recovery, multilevel modulation, and/or an inner FEC code. These effects typically cause error bursts in the pre-FEC bit stream, which may affect the FEC decoder performance either beneficially or adversely, depending on the type of code. In either case, the performance predicted based on the binary symmetric channel will be invalid. To achieve full decorrelation, the interleaver length should be equal to the burst length times the code length.

An important aspect that is often overlooked is that if the FEC code is fixed, and the rest of the system is designed for optimum performance at the corresponding FEC or GMI threshold, then the overall system will almost surely perform suboptimally [41]. A significantly better performance may be obtained by considering a different FEC OH (and hence a different threshold) and reoptimizing other system parameters. For example, there exists a fundamental trade-off between modulation order and FEC OH, which is not exploited if the code is fixed to some ad-hoc value (such as 6.7%). It is quite possible that the overall throughput is increased if the OH is increased (from say 6.7% to 20% or even higher), since

the improved error-correction capability may enable a higher-order modulation format. Similar trade-offs exist between the OH and other important system parameters, such as baudrates, WDM spacing, DSP algorithms, amplifier spacing, launch powers, and many others.

IV. DISCUSSION AND RECOMMENDATIONS

We have introduced some information-theoretic concepts and tools useful for the design and analysis of optical fiber systems. The widespread notion of a so-called nonlinear Shannon limit for optical fibers and its more rigorous interpretation as an AIR with mismatched decoding have been discussed. The problem of predicting the post-FEC BER from pre-FEC metrics has been also considered, discussing the use and validity of different predictors for HD and SD decoding, such as the pre-FEC BER and the GMI. To conclude, we give the following recommendations to optical communications engineers.

- When the channel statistics are unknown or too complicated, as for nonlinear optical fibers, use the AIR with mismatched decoding as a figure of merit.
- Consider the nonlinear Shannon limit just as a capacity lower bound, achievable by systems optimized for the AWGN channel.
- Keep looking for improved transmission and detection strategies, as the capacity limit of optical fiber systems is still unknown.
- Use Gaussian models and approximations with caution: they provide accurate results for conventional systems, but may lead to overlooking potential AIR gains and underestimating channel capacity.
- To predict the post-FEC performance of systems with HD-FEC decoding, use the pre-FEC BER and compare it with HD-FEC thresholds whose origin and correctness can be validated in references. Do not copy someone

else's reference for a threshold without checking that the claimed threshold is actually there.

- Use the pre-FEC GMI to predict the post-FEC performance of systems with binary SD-FEC decoding.
- Consider multiple thresholds to optimize the FEC OH (code rate), which may significantly improve the system performance.
- Use (or assume) a large interleaver—otherwise, post-FEC performance predictions based on FEC thresholds or other pre-FEC metrics may be very inaccurate.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3/4, pp. 379–423/623–656, July/Oct. 1948.

[2] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[3] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. San Diego, CA: Academic Press, 2001.

[4] E. Wang and C. R. Menyuk, "Polarization evolution due to the Kerr nonlinearity and chromatic dispersion," *J. Lightw. Technol.*, vol. 17, no. 12, pp. 2520–2529, Dec. 1999.

[5] 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.

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

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.

[8] M. Franceschini, G. Bongiorni, G. Ferrari, R. Raheli, F. Meli, and A. Castoldi, "Fundamental limits of electronic signal processing in direct-detection optical communications," *J. Lightw. Technol.*, vol. 25, no. 7, pp. 1742–1753, July 2007.

[9] G. Colavolpe, T. Foggi, A. Modenini, and A. Piemontese, "Faster-than-Nyquist and beyond: how to improve spectral efficiency by accepting interference," *Opt. Exp.*, vol. 19, no. 27, pp. 26 600–26 609, Dec. 2011.

[10] M. Secondini, T. Foggi, F. Fresi, G. Meloni, F. Cavaliere, G. Colavolpe, E. Forestieri, L. Potí, R. Sabella, and G. Prati, "Optical time–frequency packing: Principles, design, implementation, and experimental demonstration," *J. Lightw. Technol.*, vol. 33, no. 17, pp. 3558–3570, Sept. 2015.

[11] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3755–3763, Nov. 2005.

[12] P. Serena, A. Bononi, and G. Colavolpe, "On the nonlinear capacity with memory of PS-QPSK and PDM-QPSK in WDM non-dispersion managed links," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2012, paper We.2.C.1.

[13] 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.

[14] 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.

[15] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 24, no. 30, pp. 3875–3879, Dec. 2012.

[16] A. Splett, C. Kurtzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, vol. 2, 1993, pp. 41–44.

[17] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, June 2001.

[18] A. D. Ellis, Z. Jian, and D. Cotter, "Approaching the non-linear Shannon limit," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 423–433, Feb. 2010.

[19] M. Secondini and E. Forestieri, "Scope and limitations of the nonlinear Shannon limit," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 893–902, Feb. 2017.

[20] D. Marsella, M. Secondini, E. Agrell, and E. Forestieri, "A simple strategy for mitigating XPM in nonlinear WDM optical systems," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2015, p. Th4D.3.

[21] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *J. Lightw. Technol.*, vol. 34, no. 16, pp. 3884–3889, Aug. 2016.

[22] M. Sorokina, S. Sygletos, and S. Turitsyn, "Ripple distribution for nonlinear fiber-optic channels," *Opt. Exp.*, vol. 25, no. 3, pp. 2228–2238, Feb. 2017.

[23] E. Agrell, "Conditions for a monotonic channel capacity," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 738–748, Mar. 2015.

[24] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Phys. Rev. Lett.*, vol. 91, no. 20, Nov. 2003, paper 203901.

[25] E. Agrell and M. Karlsson, "Influence of behavioral models on multiuser channel capacity," *J. Lightw. Technol.*, vol. 33, no. 17, pp. 3507–3515, Sep. 2015.

[26] IEEE Computer Society, "IEEE standard for Ethernet," IEEE Std 802.3-2015, [Online] http://doi.org/10.1109/IEEESTD.2016.7428776, 2016.

[27] ITU-T, "Forward error correction for submarine systems," [Online] http://www.itu.int/rec/T-REC-G.975, 2000.

[28] ——, "Forward error correction for high bit-rate DWDM submarine systems," [Online] http://www.itu.int/rec/T-REC-G.975.1, 2004.

[29] K. Roberts, "Deployable optical systems at 40Gb/s and beyond," in *Proc. SPIE*, ser. Optical Transmission, Switching, and Subsystems V, 67830B, vol. 6783, 2007.

[30] H. Sun, K. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Optics Express*, vol. 16, no. 2, pp. 873–879, Jan. 2008.

[31] L. E. Nelson, S. L. Woodward, S. Foo, X. Zhou, M. D. Feuer, D. Hanson, D. McGhan, H. Sun, M. Moyer, M. O. Sullivan, and P. D. Magill, "Performance of a 46-Gbps dual-polarization QPSK transceiver with real-time coherent equalization over high PMD fiber," *J. Lightw. Technol.*, vol. 27, no. 3, pp. 158–167, Feb. 2009.

[32] K. Roberts, Q. Zhuge, I. Monga, S. Gareau, and C. Laperle, "Beyond 100 Gb/s: capacity, flexibility, and network optimization," *J. Opt. Commun. Netw.*, vol. 9, no. 4, pp. C12–C24, Apr. 2017.

[33] Y. Miyata, K. Kubo, K. Onohara, W. Matsumoto, H. Yoshida, and T. Mizuochi, "UEP-BCH product code based hard-decision FEC for 100 Gb/s optical transport networks," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2012.

[34] M. Scholten, T. Coe, and J. Dillard, "Continuously-interleaved BCH (CI-BCH) FEC delivers best in class NECG for 40G and 100G metro applications," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2010.

[35] 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," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2012.

[36] L. M. Zhang and F. R. Kschischang, "Staircase codes with 6% to 33% overhead," *J. Lightw. Technol.*, vol. 32, no. 10, pp. 1999–2002, May 2014.

[37] 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.

[38] 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.

[39] L. Szczecinski and A. Alvarado, *BitInterleaved Coded Modulation: Fundamentals, Analysis and Design*. Wiley, 2015.

[40] 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.

[41] B. P. Smith and F. R. Kschischang, "Future prospects for FEC in fiber-optic communications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1245–1257, Sept./Oct. 2010.