

Capacity Limits of Optical Fiber Networks

René-Jean Essiambre, *Senior Member, IEEE, Fellow, OSA*, Gerhard Kramer, *Fellow, IEEE*, Peter J. Winzer, *Fellow, IEEE*, Gerard J. Foschini, *Fellow, IEEE*, and Bernhard Goebel, *Student Member, IEEE*

(Invited Paper)

Abstract—We describe a method to estimate the capacity limit of fiber-optic communication systems (or “fiber channels”) based on information theory. This paper is divided into two parts. Part 1 reviews fundamental concepts of digital communications and information theory. We treat digitization and modulation followed by information theory for channels both without and with memory. We provide explicit relationships between the commonly used signal-to-noise ratio and the optical signal-to-noise ratio. We further evaluate the performance of modulation constellations such as quadrature-amplitude modulation, combinations of amplitude-shift keying and phase-shift keying, exotic constellations, and concentric rings for an additive white Gaussian noise channel using coherent detection. Part 2 is devoted specifically to the “fiber channel.” We review the physical phenomena present in transmission over optical fiber networks, including sources of noise, the need for optical filtering in optically-routed networks, and, most critically, the presence of fiber Kerr nonlinearity. We describe various transmission scenarios and impairment mitigation techniques, and define a fiber channel deemed to be the most relevant for communication over optically-routed networks. We proceed to evaluate a capacity limit estimate for this fiber channel using ring constellations. Several scenarios are considered, including uniform and optimized ring constellations, different fiber dispersion maps, and varying transmission distances. We further present evidences that point to the physical origin of the fiber capacity limitations and provide a comparison of recent record experiments with our capacity limit estimation.

Index Terms—, Amplified spontaneous emission, Brillouin scattering, channel coding, detection, fiber nonlinearity, information rates, information theory, modulation, noise, optical networks, Raman scattering.

PART 1: DIGITIZATION, CAPACITY, AND CONSTELLATIONS

I. INTRODUCTION

DETERMINING an ultimate limit to the rate at which one can reliably transmit information over a physical medium in a given environment is an endeavor having both fundamental

Manuscript received November 25, 2009. Current version published February 24, 2010. This work was supported by the Defense Advanced Research Projects Agency under Grant HR0011-06-C-0098.

R.-J. Essiambre, G. J. Foschini, and P. J. Winzer are with Bell Laboratories, Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: rjessiam@alcatel-lucent.com).

G. Kramer was with Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ 07974 USA. He is now with the Department of Electrical Engineering, University of Southern California, Los Angeles CA 90089-2565 USA.

B. Goebel is with the Institute for Communications Engineering (LNT), Technische Universität München, D-80290 Munich, Germany.

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.2009.2039464

and practical interests [1]–[4]. Such a limit is referred to as the *channel capacity* and the process of evaluating this limit leads to a better understanding of the technologies needed to approach it.

Capacity evaluations require information theory [1] that must be adapted to the specific characteristics of the channel under study. A pragmatic approach is to define the channel as “that part of a communication system that the designer is unable or unwilling to change” [5], [6]. Using this approach, capacity evaluations have been performed for a variety of physical media such as twisted-pair copper cables [7]–[10], coaxial cables [9], [11], wireless [12]–[15], and satellite communications [16]–[18], respectively. The goal of Part 1 of this paper is to introduce basic concepts for evaluating channel capacities. We refer to these tools in the second part of the paper where we evaluate the capacity of the “fiber channel.”

II. DIGITIZATION

A. Spectra and Sampling

Information is usually transmitted using electromagnetic waves over a physical medium (copper wires, coaxial cable, atmosphere, space, etc.). One often represents such waves by real, analog signals $x(t)$ referred to as *waveforms*. Suppose $x(t)$ is *bandlimited* to W Hz, i.e., the support of its Fourier transform, or spectrum, $X(f)$ is within the frequency set $\{f : |f| < W\}$, as shown in Fig. 1(a). We can then represent $x(t)$ by regularly spaced signal samples $\{x_k\}$ taken at the *Nyquist rate* of $2W$ samples per second [1],[19]–[21]. The signal may be reconstructed from the samples by multiplying the samples by a sinc function defined as $\text{sinc}(t) = \sin(\pi t)/(\pi t)$ for $t \neq 0$ and $\text{sinc}(0) = 1$, i.e., we have (see Fig. 2)

$$x(t) = \sum_k x_k \text{sinc}(2Wt - k). \quad (1)$$

Another representation of the real signal $x(t)$ follows because its spectrum satisfies $X(-f) = X^*(f)$ where a^* is the complex conjugate of a . In other words, the negative frequency components of $X(f)$ are redundant for real signals. Thus, we may represent $x(t)$ by using $X(f)$ at positive frequencies only, as shown in Fig. 1(b). This spectrum is referred to as the *single sideband* (SSB) version of $X(f)$.

Suppose next that we are interested in a *passband* signal [22]–[25]

$$x(t) = \sqrt{2}x_c(t)\cos(2\pi f_c t) - \sqrt{2}x_s(t)\sin(2\pi f_c t) \quad (2)$$

where $f_c > W/2$ is the carrier frequency and $x_c(t)$ and $x_s(t)$ are bandlimited to $W/2$ Hz. The factor $\sqrt{2}$ normalizes the energy

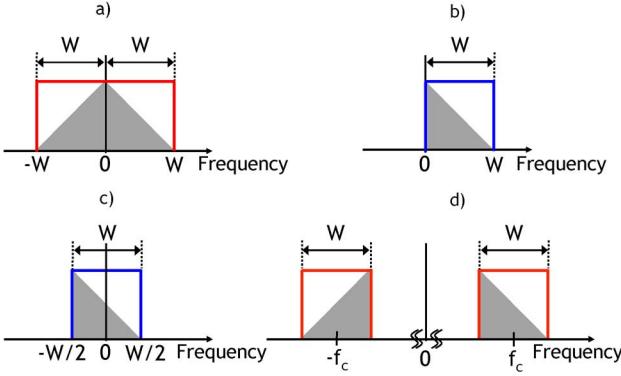


Fig. 1. Spectral amplitudes of baseband and passband signals. (a) Real baseband signal spectrum. (b) Complex signal spectrum (single-sideband spectrum, real and imaginary parts). (c) Complex baseband signal spectrum (real and imaginary parts). (d) Real passband signal spectrum (cosine and sine parts).

of the $\cos(\cdot)$ and $\sin(\cdot)$ functions. We may represent (2) by defining a *complex baseband signal*

$$x_b(t) = x_c(t) + ix_s(t) \quad (3)$$

and writing

$$x(t) = \sqrt{2}\Re\{x_b(t)e^{i2\pi f_c t}\} \quad (4)$$

where $\Re\{a\}$ is the real part of a and $i = \sqrt{-1}$. The real signals $x_c(t)$ and $x_s(t)$ are sometimes referred to as in-phase (I) and quadrature (Q) components. The signal $x_b(t)$ is sometimes referred to as the *complex envelope* of $x(t)$.

The complex signal $x_b(t)$ has spectral support $\{f : |f| < W/2\}$ as shown in Fig. 1(c), while the signal $x(t)$ has spectral support $\{f : f_c - W/2 < |f| < f_c + W/2\}$ as shown in Fig. 1(d). We again say that $x(t)$ has bandwidth W Hz. We can sample $x_c(t)$ and $x_s(t)$ at their Nyquist rate of W samples per second and these samples $\{x_{c,k}\}$ and $\{x_{s,k}\}$ may be represented as *complex samples* $\{x_k = x_{c,k} + ix_{s,k}\}$ of the signal $x_b(t)$. The signal $x_b(t)$ is reconstructed from the complex samples as follows [19, see Fig. 2], [13, Ch. 2], [26, Ch. 6], and [27, Ch. 8]

$$x_b(t) = \sum_k x_k \text{sinc}(Wt - k). \quad (5)$$

The reader might wonder why (1) and (5) have the sampling rates $2W$ and W , respectively. In fact, the sampling in (5) also represents a sampling rate of $2W$ *real* samples per second since the x_k in (5) are *complex* numbers. Mathematically, the passband signal of (2) is projected onto the signals

$$\sqrt{2} \cos(2\pi f_c t) \text{sinc}(Wt - k)$$

and

$$-\sqrt{2} \sin(2\pi f_c t) \text{sinc}(Wt - k)$$

$k = \dots, -1, 0, 1, 2, 3, \dots$, rather than the signals $\text{sinc}(2Wt - k)$ as is done for passband signals [13, Ch. 2], [26, Ch. 6], Ch. 6, and [27, Ch. 7]. The difference in the representations of (1) and (5) can be seen by considering the baseband signal of (1) to be a passband signal at the carrier frequency $f_c = W/2$, assuming

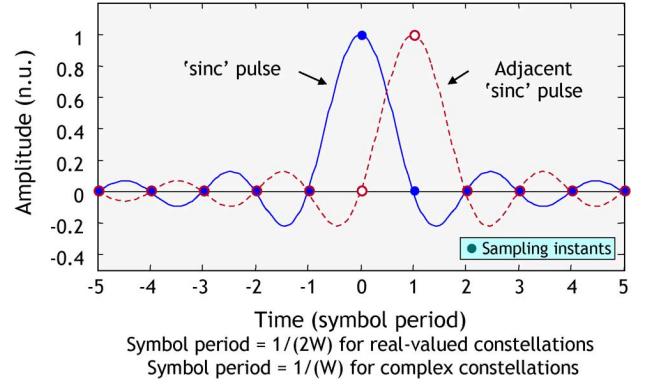


Fig. 2. Two adjacent sinc pulses. The value of a sinc pulse at regular sampling instants is zero except at one sampling instant for each pulse. This property makes sinc pulses free from intersymbol interference (ISI) between symbols at these sampling instants.

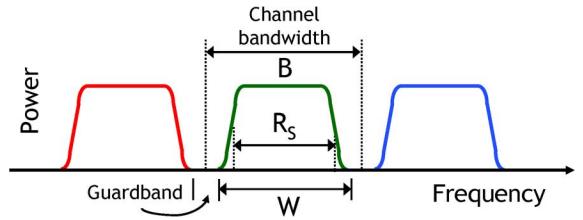


Fig. 3. Symbol rate R_s , spectral support W , and wavelength-division multiplexing (WDM) channel bandwidth B for a transmit pulse with a square-root raised-cosine spectrum.

now that $f_c = W/2$ is permitted. In both representations (1) and (5), a sampling rate of $2W$ real samples per signal is necessary and sufficient.

One sometimes encounters an alternative way of viewing passband signals, related to SSB modulation. For instance, we can generate a passband signal with spectrum shown in Fig. 1(d) corresponding to a real-valued baseband signal $x(t)$ with spectrum shown in Fig. 1(a) by using the following steps.

- 1) Strip off one sideband of $x(t)$ by using a *Hilbert filter* [26, p. 200], [27, Ch. 7], Ch. 7 to generate a complex signal with spectral support shown in Fig. 1(b). Now frequency up-convert the signal by $f_c - W/2$ Hz and transmit the real part of the resulting signal.
- 2) Modulate $x(t)$ with the carrier $\cos(2\pi(f_c - W/2)t)$ and then eliminate one sideband by using a bandpass filter.

We observe that SSB modulation, being effectively an alternative way of generating passband signals, does not gain capacity over dual sideband modulation with complex baseband signals, as described above.

We remark that we distinguish between the signal spectral support W , the frequency bandwidth B assigned to the signal within the optical network, and the symbol rate R_s at which one is modulating the transmit pulse. The symbol rate is $R_s = 1/T$, where T is the symbol period. For example, consider Fig. 3 and suppose we use passband communication, a transmit pulse with a square-root raised-cosine spectrum [25], [23] and a roll-off factor $\rho = 0.3$, and B that is 20% larger than W . In this case, we have $W = (1 + 0.3)R_s$ and $B = (1 + 0.2)W$. In general, we consider $R_s \leq W \leq B$.

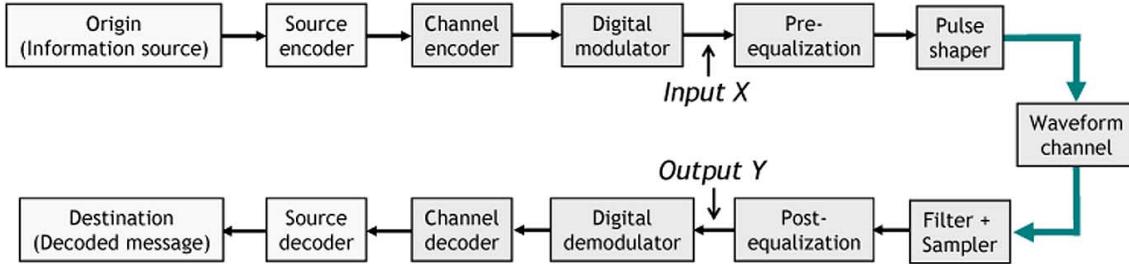


Fig. 4. Schematic representation of the main functions of a digital communication channel. The lighter boxes represent the source compression and decompression functions while the darker boxes represent the communication channel and the related coding and modulation functions. The thicker arrows represent analog waveforms.

B. Transmitter: Digital-to-Analog

The *discrete-time* signal representations described earlier are eminently practical: to generate a passband signal $x(t)$ that is bandlimited to W Hz one may generate R_s complex symbols per second, multiply each symbol by a *pulse* that is bandlimited to W Hz, and transmit these pulses in sequence. In other words, signal generation may be separated into two distinct parts: *modulation* defined by a discrete set of values called the *modulation alphabet* or *constellation*, and *pulse shaping* to create the pulse waveforms [3]. The size of the constellation determines the maximum information that each symbol can carry, while pulse shaping affects the spectral width occupied by the signal. The pulse shaping can be synthesized digitally using a digital-to-analog converter (DAC) [28], [29]. Constellations are treated in Section IV, so we consider pulse shaping next.

The *pulse shape* may be chosen so that there is no ISI between successive symbols. For instance, the sinc pulse in Fig. 2 is zero for all sampling instants $t = kT$ except for $k = 0$, where T is the symbol period. However, the sinc pulse amplitude decays slowly ($\sim 1/|t|$) so that there is significant ISI if the sampling is imperfect. Another commonly used pulse shape has the *raised-cosine* spectrum [26, Ch. 6]

$$S(f) = \begin{cases} T, & |f| \leq \frac{1-\rho}{2T} \\ T \cos^2 \left[\frac{\pi T}{2\rho} \left(|f| - \frac{1-\rho}{2T} \right) \right], & \frac{1-\rho}{2T} < |f| < \frac{1+\rho}{2T} \\ 0, & \text{else} \end{cases} \quad (6)$$

where ρ is the *roll-off* factor. The pulse shape corresponding to (6) is

$$s(t) = \text{sinc}(t/T) \frac{\cos(\pi\rho t/T)}{1 - 4\rho^2 t^2/T^2}, t \neq 0. \quad (7)$$

Clearly, we have $s(kT) = 0$ for $k \neq 0$ so there is no ISI. Moreover, the pulse decays much more quickly ($\sim 1/|t|^3$) than the sinc pulse if $\rho > 0$, thereby significantly reducing ISI with imperfect sampling. The price paid is that the spectrum has a larger bandwidth $W = R_s(1+\rho)$ than the sinc pulse. The choice $\rho = 0$ recovers the sinc pulse.

A third commonly used pulse has the *square-root raised-cosine* spectrum $\sqrt{S(f)}$ with $S(f)$ given by (6). The corresponding time-domain pulse is *not* zero at the sampling times $\pm T$ and, therefore, exhibits ISI. The reason for using this pulse shape is that by placing a square-root raised-cosine spectrum

filter at both the transmitter and receiver, the overall pulse shape has a raised-cosine spectrum [15, Ch. 11]. Furthermore, the receiver filter now acts as a *matched filter* for the transmit pulse. We use the matched filter receiver because it maximizes the signal-to-noise ratio (SNR) (see Section III-C) for channels with additive noise. However, note that the nonlinear fiber channel studied in Part 2 of this paper may have a different optimum receiver structure, which is a topic of further investigations. We refer to [26, Ch. 6], [14, Ch. 5], and [15, Ch. 11] for more discussion on pulse shapes and matched filters.

Finally, we remark that the signal spectrum can be “narrowed” below the minimum bandwidth associated to the symbol rate R_s by using *correlative* methods such as partial-response signaling [30], [31] or continuous phase modulation (CPM) [32], [33]. Correlative methods introduce *memory*. For instance, the duobinary signaling in [34, Fig. 6] modulates a pulse twice as fast as usual, perhaps resulting in $R_s > W$, and thereby introducing memory into successive Nyquist-rate symbols while doubling the data rate in a given spectral band [34]. However, as was pointed out in [35], such modulations are better viewed as an *encoding operation* followed by a *memoryless* modulator. In other words, correlative methods such as duobinary signaling or CPM are better viewed as *coded* versions of the usual signaling.

For example, the duobinary signaling in [23] and [24] with a pulse shape $s(t)$ is equivalent to a unit-memory, rate 1, digital encoder followed by a memoryless modulator with a constellation size of 3 and a pulse shape $s(t)$.¹ A closely related CPM method is known as minimum-shift keying (MSK) [36], and it can also be represented as a unit-memory, rate 1, digital encoder followed by a memoryless modulator and a pulse-shaper [37], [38]. We can, therefore, focus our attention on finding the capacity of signaling with memoryless modulators and the usual pulse shaping.

C. Receiver: Analog-to-Digital

At a receiver we may capture all the information in a noisy bandlimited signal $y(t)$ by filtering the signal to reject noise and interference outside the band of interest, and subsequently *sampling* the signal at its Nyquist rate. As an additional practical step, one usually *quantizes* the amplitude of the signal samples

¹See Section III and Fig. 4 for the meaning of “unit-memory,” “rate 1,” and “digital encoder.” The encoder for duobinary signaling shapes the spectrum and this type of coding is called *line coding*.

at detection to a discrete and finite set of values that are represented by sampling bits [23]. This is done by an analog-to-digital converter (ADC) [39], [29].

Consider the transmitted waveform $x(t)$ and suppose that each symbol x_k in (5) takes on one of M complex values. The combination of an ADC and demodulator that puts out more than M values per sample is called a *soft-decision* detector and it leads to two scenarios for the digital demodulator and subsequent decoder shown in Fig. 4 (see [26, Ch. 8], [14, Ch. 8], and [27, Ch. 29]). In the first scenario, called *hard-decision* decoding, the demodulator decides which modulation symbol was transmitted and passes its decision to the decoder (see Fig. 4); the decoder operates on these *hard* decisions. In the second scenario, called *soft-decision* decoding, some or all of the sampling bits are passed to the decoder and the decoder uses this *soft* information to decode [40]. In other words, the digital demodulator in Fig. 4 is effectively removed. Obviously, using soft-decision decoding with many quantization levels is preferable for performance, while using hard-decision decoding with few quantization levels reduces complexity. A summary of the performance and challenges of high-speed ADCs can be found in [41]–[43].

As we shall see shortly, using discrete-time and discrete-alphabet signals makes sense at both the transmitter and the receiver because noise limits our ability to extract information. The process of converting a continuous-time and amplitude signal $y(t)$ to a discrete-time and discrete-amplitude-and-phase signal $\{y_k\}$ is referred to as *digitization*. One of the many key insights provided by Shannon's information theory is that it suffices to consider digitized signals to approach the ultimate capacity limits of noisy channels [1]. Shannon's work is generally recognized as having given birth to digital communications and laying the foundation of today's "computer and information age".

III. INFORMATION THEORY

The process of transmitting information between an information source and a receiver can be represented by the basic building blocks shown in Fig. 4 [1], [25]. The process can be divided into two sets of operations: source encoding and decoding (light boxes in Fig. 4) [44] and the channel with its encoder, modulator, pre-equalizer, pulse shaper and inverse functions (dark boxes in Fig. 4) [2], [3].² The process of source encoding can be loosely described as a process to remove redundancy in the information source so as to produce a purely random data source. In this paper, we assume that the data are already compressed and can, therefore, be represented by a sequence of independent and identically distributed (i.i.d.) bits each of which are equally likely to be 0 and 1. Shannon showed in [1] that separating source and channel coding as shown in Fig. 4 incurs no loss in communication rate.

Consider, therefore, the channel encoder and decoder. The channel encoder takes as input a stream of bits and puts out a stream of bits with added *redundancy*, usually in the form of *parity checks* [40]. The encoder rate \tilde{R}_c represents the number

²Line coding for spectral shaping and security coding for secrecy and authentication are other types of coding. Line coding was treated briefly in Section II-B; we consider neither source nor security coding.

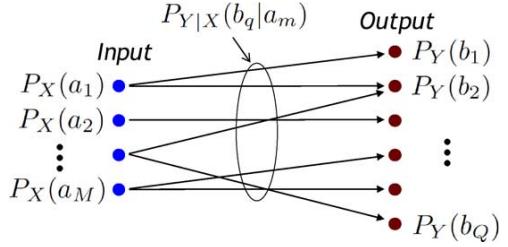


Fig. 5. Schematic representation of the channel probability distributions.

of input bits per output bit³ and satisfies $0 < \tilde{R}_c \leq 1$. $\tilde{R}_c = 1$ corresponds to uncoded data. The *code overhead* is defined as $(1 - \tilde{R}_c)/\tilde{R}_c$ and is usually given as a percentage. For instance, a code with rate $\tilde{R}_c = 15/16$ has a 6.7% overhead. The *encoder memory* refers to the number of successive input bits, in excess of a single bit, required to compute each output bit. For example, a *unit-memory* encoder requires two successive input bits to compute each output bit.⁴

The encoder's output bits are mapped onto a discrete set of values from a *modulation constellation* of size M . The number of (coded) bits per modulation symbol is, therefore, $\log_2(M)$, and the number of information bits per modulation symbol is $\tilde{R} = \tilde{R}_c \log_2(M)$. This combination, or any other combination, of channel encoding and modulation is referred to as *coded modulation*.

The pre- and post-equalization functions serve as "generalized filters" that can remove channel memory or perform non-linear operations, such as described in Section X-D.

The receiver quantizes each received pulse onto a (perhaps large) discrete set of values, as described in Section II-C. The *channel* from the digital modulator's output X to the digital demodulator's input Y is, therefore, a discrete-input, discrete-output channel. For example, Fig. 5 depicts a channel with modulation constellation $\mathcal{X} = \{a_1, a_2, \dots, a_M\}$, output alphabet $\mathcal{Y} = \{b_1, b_2, \dots, b_Q\}$, and channel probabilities $P_{Y|X}(b_q|a_m)$, $1 \leq m \leq M$, $1 \leq q \leq Q$, i.e., $P_{Y|X}(b_q|a_m)$ is the probability of $Y = b_q$ under the condition that $X = a_m$ was sent.

We shall review the notion of channel capacity for discrete-input, discrete-output channels in the following section. However, it is insightful, and computationally useful, to consider what is possible if one were permitted to increase the sizes of the modulation and receiver sampling alphabets without limit. For this reason, we will also consider as follows:

- 1) complex-input, complex-output channels;
- 2) discrete-input, complex-output channels; and
- 3) ring-input, complex-output channels.

The motivation for considering the first of these channels is to determine what is ultimately possible without considering transmitter or receiver complexity. The motivation for considering the second channel is to see what is possible if the receiver can quantize as finely as desired. One reason for consid-

³A tilde distinguishes rates that are *not* "per second" (input bits/output bits, bits/symbol) from rates that are "per-second" (bits/second, symbols/second).

⁴More generally, an encoder can take in a stream of *symbols* and put out a stream of *symbols*. For example, a symbol might represent a block of bits. A *unit-memory* encoder might then be viewed as requiring two successive input symbols to compute each output symbol.

ering ring constellations is to later take advantage of channel rotational symmetries that simplify capacity computations (see Section X-C). Finally, in what follows we also discuss channels with memory to motivate and justify our approach for finding a capacity lower bound estimate in Part 2 of this paper.

A. Discrete-Input, Discrete-Output Channels

Consider the discrete-time, discrete-input, discrete-output channel depicted in Fig. 5. The channel takes as input a sequence x_1, x_2, \dots, x_n and puts out a sequence y_1, y_2, \dots, y_n , where y_k is a noisy function of x_k , $k = 1, 2, \dots, n$. Every x_k and y_k takes on a value in \mathcal{X} and \mathcal{Y} , respectively. The noise is described by the conditional probability distribution $P_{Y|X}(\cdot)$ that is time-invariant, i.e., the probability that the input x_k produces the output y_k is $P_{Y|X}(y_k|x_k)$ for all k . Following [2], we shall usually refer to random variables with upper case letters X, Y, X_k, Y_k and realizations of these variables using lowercase letters x, y, x_k, y_k . For notational convenience, we shall sometimes use the sequence notation $x^k = x_1, x_2, \dots, x_k$ and similarly $y^k = y_1, y_2, \dots, y_k$.

To describe capacity, it is useful to define the *entropy* of the input random variable X as (see [1]–[3])

$$H(X) = \sum_a -P_X(a) \log_2 P_X(a) \quad (8)$$

where $P_X(a)$ is the probability that the random variable X takes the value a . Symbols a with $P_X(a) = 0$ are not included in the summation and we follow this convention below. We have further taken the logarithm to the base 2 so that our units are in binary digits or *bits*. If the natural logarithm is used, then the units are called *nats*. One nat is $\log_2(e) = 1/\ln(2) \approx 1.4427$ bits. We remark that we can alternatively write (8) as

$$H(X) = \mathcal{E}[-\log_2 P_X(X)] \quad (9)$$

where $\mathcal{E}[\cdot]$ is the expectation operator [3, Ch. 2].

The entropy can be considered to be the *uncertainty* about X . Similarly, the uncertainty of one random variable X relative to the realization of another random variable Y can be captured by introducing a quantity called *conditional entropy*. For $Y = b$ with $P_Y(b) > 0$, conditional entropy is defined as

$$H(X|Y = b) = \sum_a -P_{X|Y}(a|b) \log_2 P_{X|Y}(a|b) \quad (10)$$

where $P_{X|Y}(a|b)$ is the probability that $X = a$ was transmitted given that we observe $Y = b$ at the receiver. Using Bayes' rule [3], [45], [46], we have $P_{X|Y}(a|b) = P_{XY}(a,b)/P_Y(b)$ where $P_{XY}(a,b)$ is the joint probability that $(X, Y) = (a, b)$. The *average* conditional entropy is written as

$$\begin{aligned} H(X|Y) &= \sum_b P_Y(b) H(X|Y = b) \\ &= \sum_{a,b} -P_{XY}(a,b) \log_2 P_{X|Y}(a|b). \end{aligned} \quad (11)$$

Using (8) and (11), we define a quantity that measures the information between two random variables X and Y , in the sense that it measures how much knowing Y reduces $H(X)$ to

$H(X|Y)$. This quantity is referred to as *mutual information* and is defined as [2, Ch. 2], [3, Ch. 2]

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= \sum_{a,b} P_{XY}(a,b) \log_2 \frac{P_{XY}(a,b)}{P_X(a)P_Y(b)}. \end{aligned} \quad (12)$$

Observe that, we can also write $I(X; Y) = H(Y) - H(Y|X) = I(Y; X)$, which is the reason for calling this quantity the *mutual information*. The *capacity* of a channel is the maximum mutual information, where the maximization is performed over all possible input distributions $P_X(\cdot)$, i.e., the capacity in bits per (channel input or output) symbol or “channel use” is [1], [2, Ch. 4], and [3, Ch. 7]

$$\tilde{C} = \max_{P_X(\cdot)} I(X; Y). \quad (13)$$

It turns out that *reliable* communication, i.e., communication with arbitrarily small nonzero error probability, is possible at the rate \tilde{R} bits per symbol if $\tilde{R} < \tilde{C}$ and is impossible if $\tilde{R} > \tilde{C}$ (the case $\tilde{R} = \tilde{C}$ behaves differently depending on the channel). Observe from (12) and (13) that the capacity is a *maximum entropy difference*.

B. Complex-Input, Complex-Output Channels

We next focus our attention on the *complex* additive white Gaussian noise (AWGN channel) (see Fig. 6 and [1], [2, Ch. 7], and [3, Ch. 9]). The channel input X and output Y are complex random variables with

$$Y = X + N \quad (14)$$

where $N = N_R + jN_I$ is *noise*, and N_R and N_I are independent Gaussian random variables each having zero mean and variance $\sigma_N^2/2$. In other words, for $n = n_R + in_I$ we have

$$p_N(n) = \frac{1}{\pi\sigma_N^2} e^{-(n_R^2+n_I^2)/\sigma_N^2}. \quad (15)$$

The input X and noise N are independent, and the channel conditional probability density is, therefore,

$$\begin{aligned} p_{Y|X}(y|x) &= p_N(y-x) \\ &= \frac{1}{\pi\sigma_N^2} e^{-((y_R-x_R)^2+(y_I-x_I)^2)/\sigma_N^2} \end{aligned} \quad (16)$$

where $x = x_R + ix_I$ and $y = y_R + iy_I$.

Suppose that X and Y have a joint density $p_{XY}(x,y)$. The entropies and mutual information are now defined as [3, Ch. 9]⁵

$$H(X) = \int -p_X(x) \log_2 p_X(x) dx \quad (17)$$

$$H(X|Y) = \int \int -p_{XY}(x,y) \log_2 p_{X|Y}(x|y) dx dy \quad (18)$$

$$I(X; Y) = \int \int p_{XY}(x,y) \log_2 \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)} dx dy \quad (19)$$

⁵The entropy of a continuous random variable X is called a *differential entropy* and is often represented using the notation $h(X)$ rather than $H(X)$. We here use the notation $H(X)$ for entropy and differential entropy.

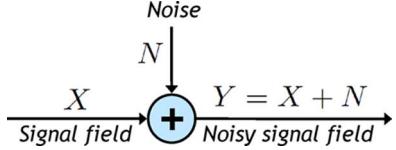


Fig. 6. Schematic representation of the AWGN channel.

assuming that all of these integrals exist. For the complex AWGN channel, we compute

$$\begin{aligned} I(X;Y) &= \int \int p_{XY}(x,y) \log_2 \frac{p_{Y|X}(y|x)}{p_Y(y)} dx dy \\ &= H(Y) - H(Y|X) \\ &= H(Y) - H(X + N|X) \\ &= H(Y) - H(N) \\ &= H(Y) - \log_2(\pi e \sigma_N^2) \end{aligned} \quad (20)$$

where the last step follows by inserting (15) into (17) with N replacing X .

Suppose we have the input constraint⁶ $\mathcal{E}[|X|^2] \leq P$ (we soon interpret P as a power). The capacity in bits per symbol is now [1], [2, Ch. 7], and [3, Ch. 9]

$$\tilde{C} = \max_{X: \mathcal{E}[|X|^2] \leq P} I(X;Y). \quad (21)$$

It is well known from Shannon's work that the optimum X has a density that is a bidimensional Gaussian distribution of the form [1], [2, Ch. 7], and [3, Ch. 9]

$$p_X(x) = \frac{1}{\pi P} e^{-(x_R^2 + x_I^2)/P}. \quad (22)$$

The resulting capacity in *bits per modulation symbol* is

$$\tilde{C} = \log_2 \left(1 + \frac{P}{\sigma_N^2} \right) \quad (23)$$

and P/σ_N^2 is referred to as an SNR.

Suppose, we use a sinc pulse (see Fig. 2) with symbol period T , bandwidth $W = 1/T$, and energy $E = PT$ Joules, respectively, for signaling and the same sinc pulse but with unit energy as a receiver (matched) filter. Suppose the noise is a Gaussian random process with a (two-sided) power spectral density of $N_0/2$ Watts/Hz across all (positive and negative) frequencies of interest [26, Sec. 7.7], [27, Sec. 25.15]. The noise power after the receiver sinc filter is, therefore, $\sigma_N^2 = N_0 W$ Watts per sample and the noise samples are independent. Using (23), the capacity in bits per second is [26, Ch. 7], [13, Ch. 5]

$$C = W \log_2 \left(1 + \frac{P}{N_0 W} \right). \quad (24)$$

⁶This constraint may be interpreted as $\mathcal{E}[|X_k|^2] \leq P$ for all k . A less stringent constraint is $\sum_{k=1}^n \mathcal{E}[|X_k|^2] \leq nP$. However, the capacity in either case turns out to be given by (21).

Alternatively, since the signal and noise *energies* are $E = PT$ and N_0 Joules/symbol, respectively, we can write (24) as

$$C = W \log_2 \left(1 + \frac{E}{N_0} \right). \quad (25)$$

Observe from (24) that $\lim_{W \rightarrow \infty} C = (P/N_0) \log_2(e)$ bits per second. In other words, capacity increases linearly with P if we can use all frequency bands. Moreover, the *spectral efficiency* SE in bits per second per Hz is [26, Ch. 7]

$$\text{SE} = \frac{C}{B} = \frac{W}{B} \log_2 \left(1 + \frac{P}{N_0 W} \right). \quad (26)$$

Suppose next that we use a pulse with energy E but at the symbol rate R_s that is less than W which is less than the WDM channel bandwidth B (see Fig. 3). For the same energy E , the signal power P is reduced by a factor of R_s/W as compared to the previous case with sinc pulses at symbol rate W . Suppose we filter the received signal by using a unit-energy matched filter before sampling. The noise energy is again N_0 by using Parseval's theorem [26, p. 115] and the noise power is $N_0 R_s$. The capacity of (24) in bits/second, thus, reduces to

$$C = R_s \log_2 \left(1 + \frac{P}{N_0 R_s} \right). \quad (27)$$

The spectral efficiency of (26) in bits/second/Hz reduces to

$$\text{SE} = \frac{C}{B} = \frac{R_s}{B} \log_2 \left(1 + \frac{P}{N_0 R_s} \right). \quad (28)$$

Note again that by keeping the energy E constant the power P in (27)–(28) is smaller than the P in (24) and (26).

C. Signal-to-Noise Ratio

We develop three SNR definitions that are commonly used. The second of these leads to an important insight based on the spectral efficiency of (28). The first definition is taken from (25) or (27), namely

$$\text{SNR} = \frac{E}{N_0} = \frac{P}{N_0 R_s}. \quad (29)$$

Equation (28) can thus be expressed as

$$\frac{C}{B} = \frac{R_s}{B} \log_2(1 + \text{SNR}). \quad (30)$$

Second, the SNR in (29) is based on the energy or power per modulation symbol. For a fair comparison among modulation formats, it is convenient to consider the SNR per information bit which we denote by SNR_b . Recall that the number of information bits per modulation symbol is $\tilde{R} = \tilde{R}_c \log_2(M)$, so we define

$$\text{SNR}_b = \frac{\text{SNR}}{\tilde{R}}. \quad (31)$$

We remark that SNR_b is often referred to as E_b/N_0 since the energy per information bit is $E_b = E/\tilde{R}$ [26, Ch. 7]. Note that both SNR and SNR_b are defined here in a single mode (i.e., single

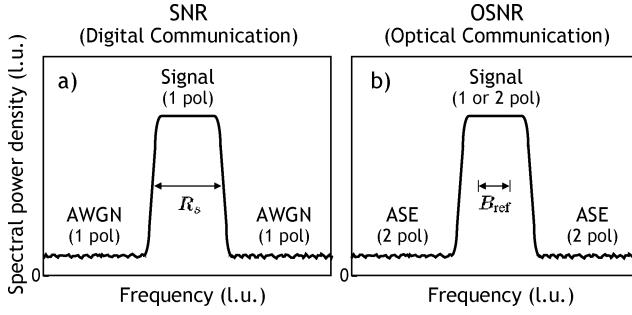


Fig. 7. Quantities involved in the definition of (a) signal-to-noise ratio (SNR) and (b) optical SNR (OSNR). l.u.: linear units, AWGN: additive white Gaussian noise, ASE: amplified spontaneous emission.

polarization) with complex signal and noise representation. Inserting (31) into (30) and using $\tilde{R}R_s \leq C$ in bits/second, we have

$$\tilde{R} \leq \log_2(1 + \tilde{R} \text{SNR}_b). \quad (32)$$

One can check that (32) has nonzero solutions only if $\text{SNR}_b > \ln(2)$. In other words, the minimum possible SNR_b , expressed in decibels (dB), is $10 \log_{10}(\ln(2)) \simeq -1.59$ dB [26, Ch. 7], [13, Ch. 5]. Note that \tilde{R} in (32) has the units of bits/symbol. The spectral efficiency in bits/second/Hz is the largest solution for \tilde{R} in (32) reduced by the factor R_s/B symbols/second/Hz. For instance, when signaling with sinc pulses without guardband we have $R_s/B = 1$ (see Fig. 3) while if we use $W = 1.3R_s$ and $B = 1.2W$ as in the example in Section II, we have $R_s/B = 1/1.2/1.3 \approx 0.641$.

Third, in optical communications one usually uses a quantity called the *optical SNR*. A schematic representation of the quantities involved in the definition of OSNR along with the corresponding quantities for SNR definitions are shown schematically in Fig. 7. The definition of OSNR is

$$\text{OSNR} = \frac{P}{2N_{\text{ASE}}B_{\text{ref}}} \quad (33)$$

where P is the total average signal power summed over the two states of polarization, N_{ASE} is the spectral density of amplified spontaneous emission (ASE, see Section IX-B1) in one polarization and the reference bandwidth B_{ref} is commonly taken to be 12.5 GHz, corresponding to a 0.1 nm resolution bandwidth of optical spectrum analyzers at 1550 nm carrier wavelength (193.4 THz carrier frequency). The factor 2 in (33) is often interpreted as accounting for both polarizations of ASE. The definition of OSNR differs from SNR by a normalization factor based on the particular choice for the *fixed* reference noise bandwidth as well as by how one accounts for signal and noise polarization modes. Using (29) and (33), one can relate SNR and OSNR directly as

$$\text{OSNR} = \frac{pR_s}{2B_{\text{ref}}} \text{SNR} \quad (34)$$

where $p = 1$ for a singly polarized signal and $p = 2$ for a polarization-multiplexed signal, and where N_{ASE} and N_0 are assumed to be equivalent (see Section IX-B1).

We remark that the information bit rate, in bits per second, is given by

$$R_b = p\tilde{R}R_s = p(\tilde{R}_c \log_2(M)) R_s \quad (35)$$

where a symbol includes the field of both polarizations. Using $\text{SNR} = \tilde{R} \text{SNR}_b$, one can, therefore, express (34) as

$$\text{OSNR} = \frac{R_b}{2B_{\text{ref}}} \text{SNR}_b. \quad (36)$$

Note that this relationship between OSNR and SNR_b depends only on the information bit rate R_b and is independent of whether one uses polarization multiplexing or not.

D. Discrete-Input, Complex-Output Channels

Sections III-B and III-C showed what is possible if we could transmit using infinitely fine modulation alphabets and infinitely fine receiver quantization. We now take a step back from these idealizations and use a discrete modulation alphabet \mathcal{X} . The mutual information of (19) becomes (see [47])

$$I(X; Y) = \sum_{a \in \mathcal{X}} P_X(a) \int p_{Y|X}(y|a) \log_2 \frac{p_{Y|X}(y|a)}{p_Y(y)} dy \quad (37)$$

where

$$p_Y(y) = \sum_{a \in \mathcal{X}} P_X(a) p_{Y|X}(y|a). \quad (38)$$

Alternatively, we can simply use (20) with (38). The capacity is thus a maximization of $H(Y)$ over $P_X(\cdot)$.

For example, suppose we use M -ary phase-shift keying (M-PSK) which means that $\mathcal{X} = \{\sqrt{P}e^{2\pi ik/M}, k = 0, 1, \dots, M-1\}$. The optimal input distribution turns out to be $P_X(\sqrt{P}e^{2\pi ik/M}) = 1/M$ for all k . This intuitively pleasing result follows by the rotational symmetry of the channel: (37) remains the same if we replace X with $Xe^{i\phi}$ for any fixed phase ϕ . Furthermore, we know that $I(X; Y)$ is concave in $P_X(\cdot)$ if $p_{Y|X}(\cdot)$ is held fixed [3, p. 33]. We can thus generate a new channel input \tilde{X} that is a probabilistic mixture of several $Xe^{i\phi}$ for different ϕ and apply Jensen's inequality [3, Sec. 2.6] to show that $I(\tilde{X}; \tilde{Y}) \geq I(X; Y)$ where \tilde{Y} is the channel output when the channel input is \tilde{X} . For M-PSK, one simply chooses a uniform mixture of $Xe^{i\phi}$ where $\phi \in \{2\pi ik/M, k = 0, 1, \dots, M-1\}$.

E. Ring-Input, Complex-Output Channels

In Part 2 of this paper, we consider X that are continuous in phase but discrete in amplitude. We refer to such X as *ring* constellations with the alphabet $\mathcal{X} = \{\sqrt{P_k}e^{i\phi}, \phi \in [0, 2\pi)\}$ for one ring or

$$\mathcal{X} = \{\sqrt{P_k}e^{i\phi}, \phi \in [0, 2\pi), k = 1, 2, \dots, r\} \quad (39)$$

for r rings for some choice of ring powers P_1, P_2, \dots, P_r . We can again use (20) to compute the mutual information and to show that the best X have a uniformly-distributed phase on every ring by using the rotational symmetry of the channel (use the arguments outlined above for M-PSK). One could further optimize the ring powers P_k , or ring amplitudes $\sqrt{P_k}, k =$

$1, 2, \dots, r$. We provide the details of the calculation of the mutual information for ring constellations in Section IV-B and in Appendix A.

There are several reasons for choosing ring constellations [48], [49]. First, such constellations approximate the Gaussian distribution as the number of rings increases if we choose the ring amplitudes $\sqrt{P_k}$ appropriately. A simple choice is to have the rings equally spaced in optical field amplitude and with equal frequencies of occupation (equal probability for choosing a transmitted symbol from any of the r rings). The ring amplitudes and occupation frequencies could both still be optimized to better approximate (22) but, as we shall see in Figs. 16 and 17, respectively, this pragmatic constellation choice already gives $I(X; Y)$ very close to the AWGN channel capacity. We treat our choice of ring constellations in more detail in Section X-C.

F. Channels With Memory

The AWGN channel is *memoryless* because the output symbol Y_k at time k depends only on the input X_k at time k , i.e., Y_k is independent of previous channel inputs X^{k-1} and outputs Y^{k-1} given X_k [3, Ch. 7]. However, many channels in practice do not fulfill these conditions and have memory. For channels with memory, the capacity can be calculated by replacing the symbols X and Y with larger and larger blocks of letters X^n and Y^n [2, Ch. 4] (recall that $X^n = X_1, X_2, \dots, X_n$ and $Y^n = Y_1, Y_2, \dots, Y_n$). For instance, if (X^n, Y^n) has the density $p_{X^n Y^n}(\cdot)$ then $I(X^n; Y^n)$ is

$$\int \int p_{X^n Y^n}(x^n, y^n) \log_2 \frac{p_{X^n Y^n}(x^n, y^n)}{p_{X^n}(x^n) p_{Y^n}(y^n)} dx^n dy^n. \quad (40)$$

Computing $I(X^n; Y^n)$ is usually very time and resource intensive, even if n is as small as 3 or 4. We remark that when the channel is memoryless, (40) normalized by n gives the same capacity as if one uses the mutual information (19) for memoryless channels.

For propagation over optical fiber considered in Part 2 of this paper, memory is introduced by fiber chromatic dispersion and fiber nonlinearity (see Section IX-D). However, our fiber channel includes reverse propagation (or back-propagation) of a single channel using digital signal processing (DSP) that removes a large portion of the channel memory (see Section X-D). Some additional memory may remain, e.g., due to imperfect channel modeling, but for simplicity we will use a memoryless model to evaluate fiber capacity in Part 2. The following paragraphs demonstrate that using such a memoryless model results in a capacity lower bound.

Suppose that the optimal input is X^n and that the corresponding output is Y^n . Let \tilde{X}^n be an input with independent and i.i.d. symbols $\tilde{X}_k, k = 1, 2, \dots, n$. Let \tilde{Y}^n be the channel output corresponding to \tilde{X}^n . The capacity is then bounded as

$$\tilde{C} = \frac{1}{n} I(X^n; Y^n) \geq \frac{1}{n} I(\tilde{X}^n; \tilde{Y}^n) \quad (41)$$

because X^n is the capacity-achieving input. We further have

$$\begin{aligned} I(\tilde{X}^n; \tilde{Y}^n) &= H(\tilde{X}^n) - H(\tilde{X}^n | \tilde{Y}^n) \\ &= \sum_{k=1}^n [H(\tilde{X}_k) - H(\tilde{X}_k | \tilde{X}^{k-1}, \tilde{Y}^n)] \end{aligned} \quad (42)$$

$$\geq \sum_{k=1}^n [H(\tilde{X}_k) - H(\tilde{X}_k | \tilde{Y}_k)] \quad (43)$$

$$= \sum_{k=1}^n I(\tilde{X}_k; \tilde{Y}_k) \quad (44)$$

where the second step follows because the \tilde{X}_k are independent and by using the chain rule for expanding entropy [3, Sec. 2.5], and the third step follows because conditioning cannot increase entropy [3, p. 29]. We now use the fact that $I(X; Y)$ is convex in $p_{Y|X}(\cdot)$ if the distribution of X is held fixed [3, p. 33]. Observe that, the $\tilde{X}_k, k = 1, 2, \dots, n$, are identically distributed with, say, the distribution of \tilde{X} . Applying Jensen's inequality [3, Sec. 2.6] to (44), we thus have

$$I(\tilde{X}^n; \tilde{Y}^n) \geq n I(\tilde{X}; \tilde{Y}) \quad (45)$$

where \tilde{Y} is the output of a channel with input \tilde{X} and where the channel density is $p_{\tilde{Y}|\tilde{X}}(\tilde{y}|\tilde{x}) = (1/n) \sum_{k=1}^n p_{\tilde{Y}_k|\tilde{X}_k}(\tilde{y}|\tilde{x})$. Inserting (45) into (41), we find that

$$\tilde{C} \geq I(\tilde{X}; \tilde{Y}). \quad (46)$$

Summarizing, we obtain a lower bound on the capacity of a channel *with memory* by using i.i.d. inputs $\tilde{X}_k, k = 1, 2, \dots, n$, and computing (say by simulation) the mutual information over the *averaged* channel. We use precisely this approach in Part 2 to compute a *fiber capacity estimate* that is a lower bound *estimate* of the actual capacity. The bound is an estimate because our calculations in Part 2 are based on particular choices for noise-models, as well as numerically obtained estimates for the parameters of these models.

IV. MODULATION CONSTELLATIONS

In Section II, we discussed analog waveforms upon which one can imprint symbols while in Section III, we discussed the information content associated with sets of symbols when optimum coding is used. This section deals with the various ways symbols can be arranged in constellations, the performance of these constellations in terms of bit-error ratio (BER) without coding, and the capacity of these constellations on the AWGN channel assuming optimum coding.

Once a pulse shape has been chosen to represent the signal [50], one can generate each symbol pulse using two oscillators in quadrature (the “sin” and “cos” oscillators described in Section II). The two oscillator outputs can be separated without introducing crosstalk since they form an orthogonal basis [22]–[25], [51]. Each symbol can be represented by a complex number, one real number for each quadrature. Since symbol values are real numbers, constellations can be continuous sets. The continuous amplitude of each quadrature is motivated by the optimality of the bidimensional Gaussian distribution for the AWGN channel (see (22) and left plot in Fig. 14). Practical considerations compel, however, to use a set of discrete constellation points (a discrete symbol alphabet).

A. Discrete Constellations

We show in Fig. 8 examples of discrete constellations using the real part of the field only while Fig. 9 shows constellations

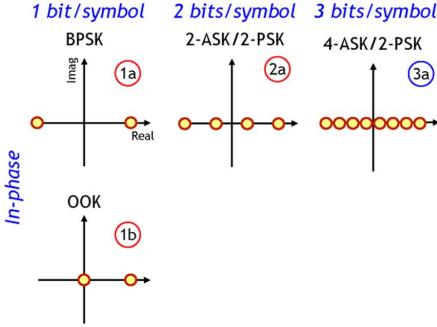


Fig. 8. Examples of constellations using only one quadrature of the field (here the real part). The number of bits/symbol is given by $\log_2(M)$ where M is the total number of symbols. The number $\log_2(M)$ of symbols is used as the first digit of the format label.

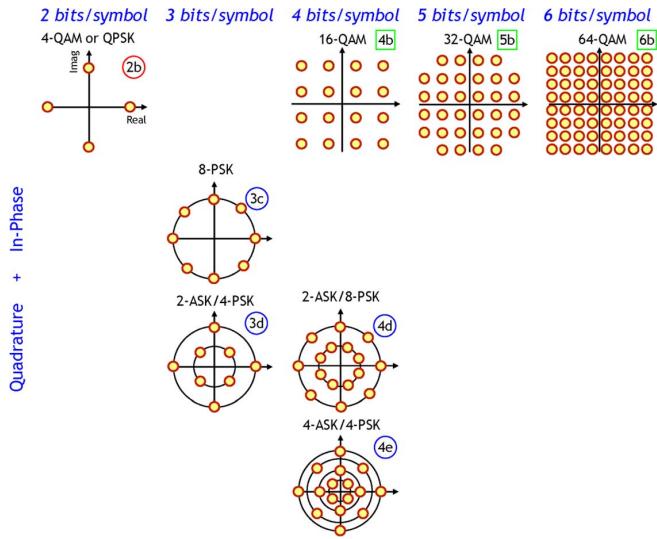


Fig. 9. Examples of constellations that use both quadratures of the field.

that utilize both field quadratures. These constellations generally carry a different number of bits per symbol, depending on the number of symbols M . A constellation can carry a maximum of $\log_2(M)$ information bits per symbol. This maximum is achieved when all the points in a constellation are used at the same frequency and in the absence of coding, which means that $\tilde{R}_c = 1$. One can design transmitters that use some constellation points more often than others, in which case different frequencies of occupation are associated with different constellation points. The conveyed information is then less than $\log_2(M)$ bits per symbol for such transmitters.

The average power associated with a constellation is given by $\mathcal{E}[|X|^2]$, i.e., the average of the square of all symbol amplitudes. This means that constellations with larger M must have their points more closely spaced together. One, therefore, expects different constellations to have different BER versus SNR curves for the AWGN channel. BER curves for the constellations of Figs. 8 and 9 are shown in Fig. 10. These BER curves apply in the absence of coding and for an identical frequency of occupation for each constellation point. Moreover, for $M \geq 2$, the bits are mapped into the constellation using *Gray mapping* (after Frank Gray, who used the term *reflected binary code* [52]) that minimizes the number of bit errors for a given symbol error

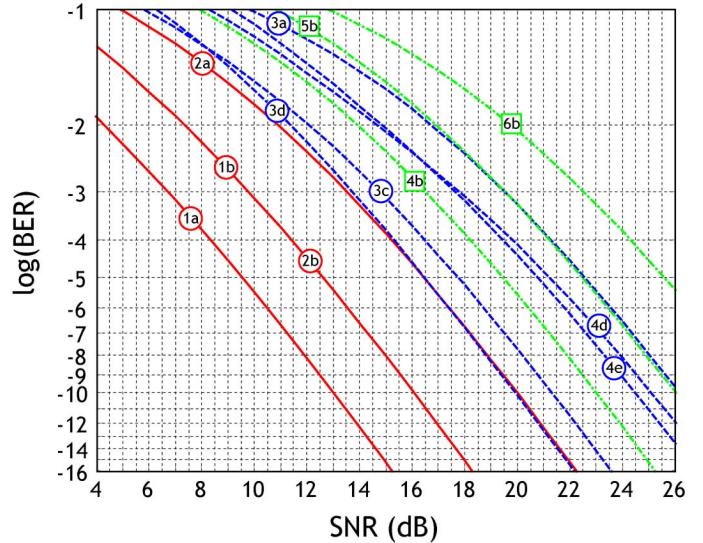


Fig. 10. BER as a function of SNR for the modulation formats of Figs. 8 and 9.

probability under the assumption that a symbol is most likely mistaken for one of its immediate neighbors [23]–[25].

The BER curve with the lowest SNR requirement in Fig. 10 represents binary phase shift keying (BPSK) (1a). The next formats having the lowest SNR requirements are the on/off keying (OOK) (1b) and 4-quadrature amplitude modulation (QAM) or quaternary phase-shift keying (QPSK) (2b) formats that have identical SNR requirements of 3 dB higher than BPSK. Therefore, between the two 1 bit/symbol formats considered, BPSK (1a) and OOK (1b), BPSK has the lowest SNR requirement by 3 dB. The next formats with low SNR requirements are 2-ASK/2-PSK (2a), where ASK stands for amplitude shift keying, and 2-ASK/4-PSK (3d) that have nearly identical SNR requirements at $\text{BER} < 10^{-4}$. The 2-bits/symbol format with the lowest SNR requirement in Fig. 10 is, therefore, 4-QAM or QPSK (2b). The 8-PSK (3c) format follows with a requirement of about 1.2 dB higher SNR than 2-ASK/4-PSK (3d) at 10^{-3} BER. The format 4-ASK/2-PSK (3a) that also supports 3 bits/symbol has a considerably higher required SNR than the two other 3 bits/symbol formats. 2-ASK/4-PSK is therefore the 3 bits/symbol format with the lowest required SNR shown here. Among the 4 bits/symbol formats (4b, 4d, and 4e), 16-QAM (4b) has the lowest SNR requirement. The QAM formats considered for 5 bits/symbol, 32-QAM (5b), and 6 bits/symbol, 64-QAM (6b), have considerably higher required SNR than lower bits/symbol formats.

The BER plot of Fig. 10 considers only the SNR and disregards the difference in the number of bits that different constellations can carry. In order to take into account the number of bits a constellation can carry, we plot the BER as a function of the SNR per bit, $\text{SNR}_b = \text{SNR}/\log_2(M)$ (see (31) with $\tilde{R}_c = 1$). The BER curves as a function of the SNR_b are shown in Fig. 11. One notices on these curves that the three pairs of formats, BPSK (1a) and 4-QAM or QPSK (2b), 2-ASK/2-PSK (2a) and 16-QAM (4b), and 4-ASK/2-PSK (3a) and 64-QAM (6b) have identical SNR per bit requirements. This can be understood by considering that the second format is an orthogonal bidimensional version of the 1-D first format. Since the two

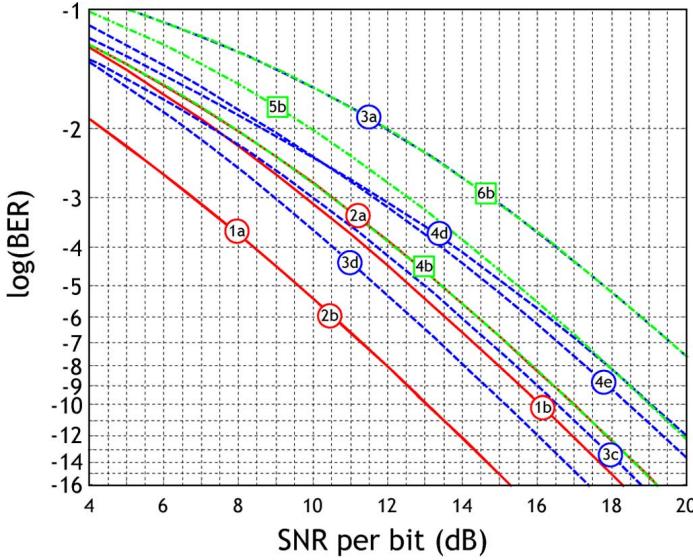


Fig. 11. BER as a function of SNR per bit for the modulation formats of Figs. 8 and 9.

field quadratures are orthogonal, modulating both quadratures conveys twice as many bits while doubling the noise variance. The SNR per bit requirements are thus identical (see [25, Sec. 4.8.4] for more explanations on BPSK and 4-QAM or QPSK). These identical requirements of SNR_b when using both quadratures is a compelling reason to make use of both quadratures for signaling. The same situation is found in optical communications when making use of polarization multiplexing: modulating both orthogonal polarizations of the optical field doubles the conveyed information but also doubles the noise power by including noise from the second polarization. As a net result, at a fixed aggregate bit rate and modulation format, the OSNR required to achieve a certain BER is the same whether one uses polarization multiplexing or not (see (36)).

The BER curves of Figs. 10 and 11, respectively, do not take into account channel coding [1]. As explained in previous sections, coding is a powerful way to reduce the BER to an arbitrarily small nonzero value. Fig. 12 shows the capacity \tilde{C} in bits/symbol (see (23)) for the formats considered so far. The Shannon capacity limit of (23) that uses a continuous bidimensional Gaussian constellation is also shown as a reference. At high SNR, all formats with a finite number of constellation points saturate to $\log_2(M)$ bits per symbol, the maximum capacity (and maximum entropy) of the respective constellation. At SNRs below the constellation capacity, one observes that constellations using both field quadratures approach their $\log_2(M)$ limits much faster than constellations that use only one quadrature (cf. constellations 1a, 1b, 2a, and 3a). For instance, one can see in Fig. 12 that 4-QAM or QPSK (2b) approach their $\log_2(M)$ limit much faster than 2-ASK/2-PSK (2a). Note also in Fig. 12 that to achieve a capacity of 3 bits/symbol, for instance, it is preferable to use *larger* constellations and code them with the appropriate (high) redundancy to achieve a low SNR requirement for the desired capacity. This is in contrast to BER curves, where larger constellations always require higher SNRs than the best smaller constellations.

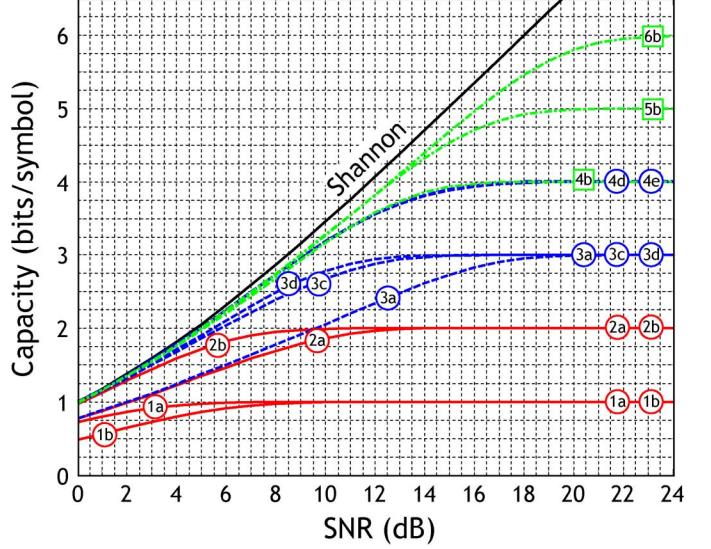


Fig. 12. Capacity as a function of SNR for the modulation formats of Figs. 8 and 9.

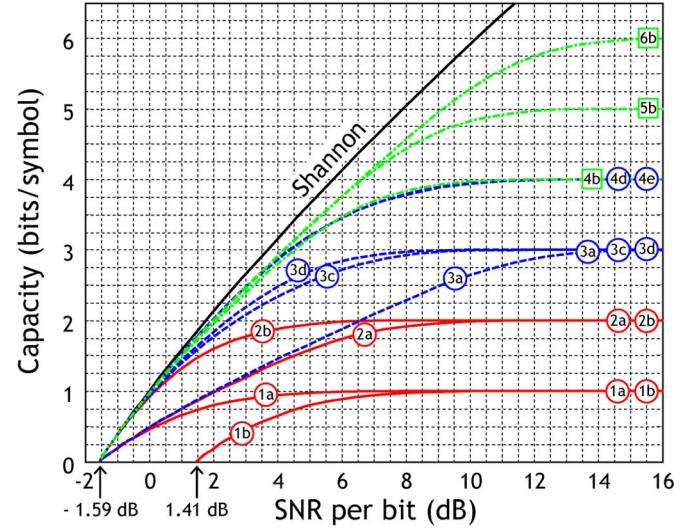


Fig. 13. Capacity as a function of SNR per bit of information for the modulation formats of Figs. 8 and 9, respectively.

The capacity as a function of SNR_b is shown in Fig. 13. It shows more clearly than in Fig. 12 the performance of different formats at low values of capacities. At low capacities ($\lesssim 0.5$ bits/symbol), all formats require an SNR per bit ($= E_b/N_0$) of $\ln(2) \approx 0.693$ or ~ -1.59 dB [47] except for the OOK format (1b) that requires 3 dB higher SNR [47]. One may note that the three other formats that use only one quadrature of the field (3a, 3c, and 3d) approach the low capacity region at half the slope of all the other formats that use both field quadratures [53]. Formats that use both field quadratures also approach the Shannon limit the closest even at higher capacities. From these capacity plots, one concludes that large constellations can be coded so as to require the minimum SNR and SNR per bit.

Finally, it is worth pointing out that all real-valued constellations and real-valued modulation waveforms could also be transmitted in their single-sideband version (see discussion along with Fig. 1). In this case, and if we were to plot the

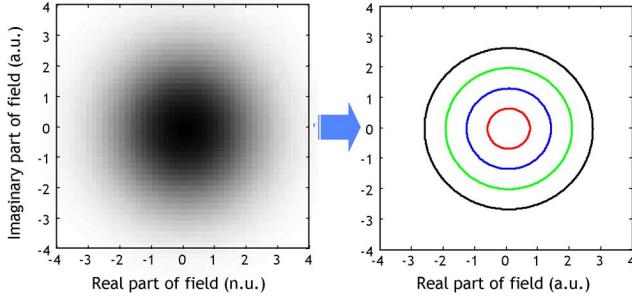


Fig. 14. Bidimensional Gaussian constellation (left plot) optimum for the AWGN channel and a four ring constellation approximation (right plot).

capacity in terms of C (in bits/s/Hz) as opposed to \tilde{C} (in bits/symbol), these formats would have the exact same limiting capacity as their complex equivalents, e.g., 2-ASK/2-PSK would have the exact same limiting capacity of 4 bits/s/Hz as 16-QAM.

We have considered in this section a number of commonly used constellations and capacity calculations assuming soft-decision decoding. In Appendix C, we consider a more extensive set of constellations, including constellations of larger sizes and different shapes. Appendix C also includes the impact of hard-decision decoding on capacity.

B. Ring Constellations

The constellations considered in Figs. 8 and 9, respectively, are used in practice because of their discrete nature that facilitates their generation. However, their capacity is limited to $\log_2 M$ bits/symbol. Such capacity limitations are alleviated by continuous constellations such as the bidimensional Gaussian constellation that leads to the Shannon capacity formula of (23). One way to understand this is that a continuous constellation allows increasing arbitrarily the “effective number” of constellation points in both quadratures. It is interesting to note that a format remains unbounded in capacity as SNR increases even if only one of the two dimensions of the constellation is continuous while the other is discretized. One way to produce a constellation not bounded in capacity is to discretize the bidimensional Gaussian constellation in *amplitude* to create concentric rings. Such a discretization allows to take advantage of *continuous* rotational symmetries of certain channels, i.e., the fact that constellation points rotated by an *arbitrary* value of phase are equivalent for these channels.

A schematic representation of the discretization in concentric rings is shown in Fig. 14. A constellation with a single ring is referred to as phase-shift keying (PSK) while constellations with two or more rings can be referred to as r-ASK/PSK, where r is the number of rings. Ring constellations having 1, 2, 4, 8, and 16 rings are shown in Fig. 15 along with their names and labels.

The capacities of the ring constellations as a function of SNR are shown in Fig. 16. We consider equidistant rings here, i.e., constellations where the radii of the outer rings are given by an integer multiple of the radius of the inner ring. The number of points on each ring is assumed to be large enough so as to be considered continuously distributed in phase, with identical frequency of occupation on each ring. The details of the ring capacity calculations are given in Appendix A.

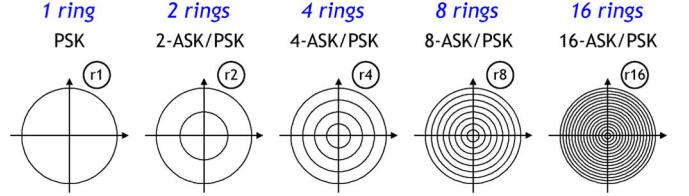


Fig. 15. Ring constellations with various numbers of rings r from 1 to 16. In ring constellations, only discrete values of amplitude are allowed while the phase can assume an arbitrary value (continuous phase). The amplitude of the outer rings is here equal to a multiple of the amplitude of the inner ring for constellations larger than one ring.

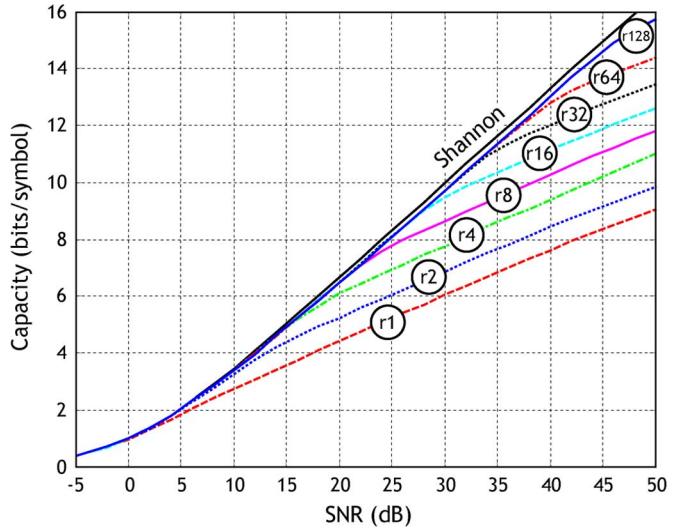


Fig. 16. Capacity as a function of SNR for the ring constellations of Fig. 15. The capacities of constellations with more than 16 rings are also presented.

The one-ring constellation starts to depart from the Shannon limit at a SNR of a few decibels, where the capacity is slightly above one bit/symbol. For more rings, departure occurs at higher SNRs and capacities. One notes that at high SNRs, the capacity for any number of rings increases at a fixed rate of ~ 0.5 bits/symbol for every doubling of the SNR in contrast to the rate of ~ 1 bit/symbol for every doubling in the SNR observed for the Shannon limit. This can be understood by the fact that, for a fixed number of rings, the capacity can only increase by increasing the number of points on the rings as the SNR increases. This confines the growth of the number of points to effectively one dimension of the constellation in contrast to the two dimensions of the bidimensional Gaussian.

We plot the capacity as a function of SNR per bit for the ring constellations in Fig. 17. As for the discrete constellations (except OOK(1b), see Fig. 13), ring constellations require a minimum of ≈ -1.59 dB per bit to transmit at any capacity. Note that the capacity curves for each ring start to change slope near a capacity of $2 \log_2(r) + 1$ bits/symbol where r is the number of rings for both Figs. 16 and 17.

Finally, we bring together the concepts presented in Part 1 by giving a numerical example in Fig. 18 of an optical field of a two-ring constellation using sinc pulses. The optical spectrum is displayed in Fig. 18(a). It is bandlimited to the bandwidth corresponding to the symbol rate R_s expressed in Hz. To facilitate visualization, a limited number of phase values on a

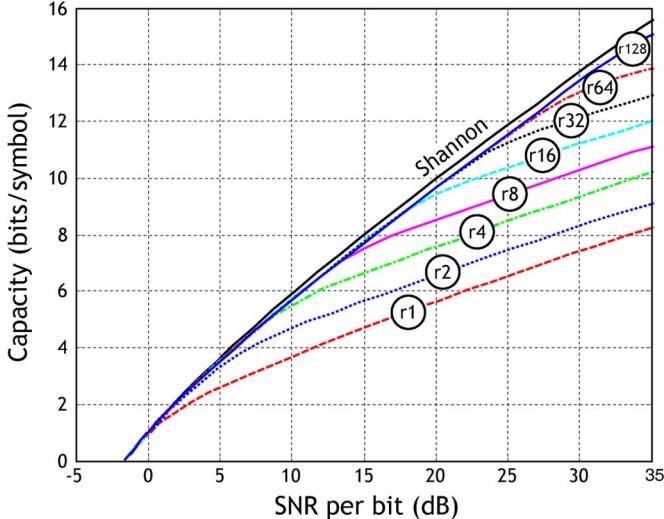


Fig. 17. Same as Fig. 16 but with capacity as a function of SNR per information bit. All constellations converge to an SNR per information bit of $10 \log_{10}(\ln(2)) \approx -1.59$ dB at low capacity.

32-point angular grid is used to represent a continuous phase (empty circles). A finer grid can be used to represent more accurately a continuous phase as needed. The constellation points actually used are shown as full circles on the angular grid. Only 32 symbols of the waveform are shown in Fig. 18(c) to facilitate visualization. The sampling instants of the time waveform of Fig. 18(c) shows where the symbols are located. Notice that in Fig. 18(c), the amplitude *between* sampling points can exceed considerably the highest value of the symbols. This phenomenon is the result of coherent addition of the sinc pulses between sampling instants and gives a “peaky” waveform, no matter how many rings are used or how many points are present in the constellation, and occurs even for binary formats. This is due to the fact that the sinc pulse extends well beyond a single symbol duration.

V. CONCLUSION

In Part 1 of this paper, we have presented basic concepts of digitization of analog waveforms, information theory, and multi-level constellations, both discrete and using ring constellations. The performance of these constellations has been evaluated in terms of BER and capacity with hard and soft decisions for the AWGN channel. These concepts, and in particular the ring constellations, will be used in Part 2 of the paper to calculate an estimate of the capacity limits of optical fibers in optically routed networks.

PART 2: FIBER CHANNEL DEFINITION AND CAPACITY ESTIMATION

VI. INTRODUCTION

The vast majority of worldwide data and voice traffic is transported using optical fibers, interconnected to form global fiber-optic networks. As the demand for bandwidth continues to increase exponentially at about 60% per year [54], it is of great

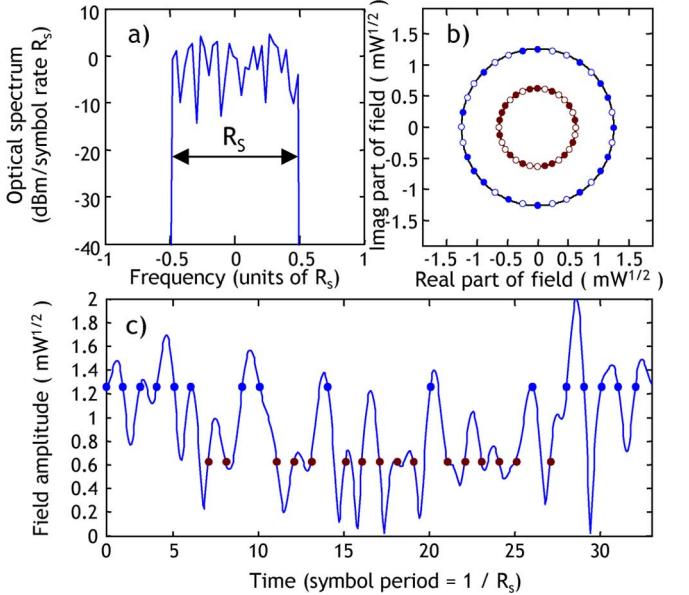


Fig. 18. Example of signal field using sinc pulses and a ring constellation. (a) Spectrum, (b) Constellation, and (c) Waveform. A small number of symbols (32) and rings (2) are represented for clarity.

interest to study the transmission capacity between two locations in such optical networks. The aim of Part 2 of this paper is to provide the most accurate capacity estimate possible for a “fiber channel” defined in the context of transporting information in *optically-routed networks* (ORNs).

Since its foundation [1], information theory has been applied to several communication channels with great success. The capacity analysis presented here is also based on information theory, with specific adaptations to the optical fiber channel. The most important difference between the optical fiber and other transmission media that have been considered for capacity analyses is the presence of *Kerr-nonlinearity*, i.e., the propagation properties of the medium change with increasing signal power. As we shall see, this property has important consequences. While linear physical media perturbed by additive noise generally result in channel capacities that increase monotonically with transmit power owing to an increasing SNR, we may find that the negative impact of nonlinear signal distortions grows at a faster rate than the SNR capacity gain at high signal powers and for a band-limited channel. This behavior may turn the channel capacity into a nonmonotonic function of the transmit power, and the channel capacity will exhibit a pronounced maximum at a given (finite) signal power level or SNR. Our fiber channel capacity estimate exhibits such behavior and therefore differs fundamentally from linear channels whose capacities have been extensively studied [5]–[18].

Applying information theory to the fiber channel faces several major challenges. An important difficulty originates from the presence of three phenomena in the fiber channel: *noise*, *filtering*, and *Kerr nonlinearity*, as visualized in Fig. 19. These three phenomena are distinct in nature, occur *simultaneously*, are *distributed* along the propagation path, and *influence* each other. Note that fiber chromatic dispersion is a form of all-pass filter and can introduce substantial memory into the channel.

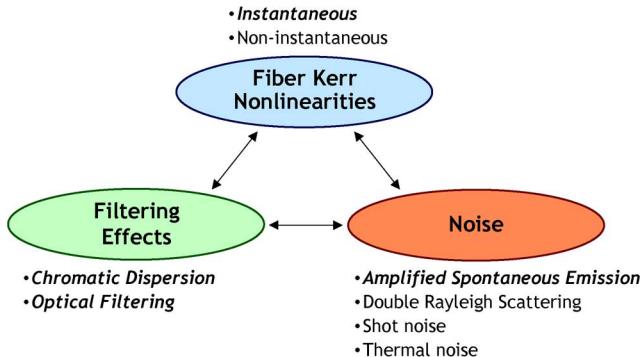


Fig. 19. List of the physical phenomena present in the optical path classified in three groups: 1) Fiber nonlinearities, 2) filtering, and 3) noise. All the phenomena mentioned in the figure are discussed in this paper. The most important phenomena limiting fiber capacity are in bold.

The various interactions between these physical phenomena may lead to *deterministic* as well as (at least partially) *stochastic* impairments.

There have been previous studies on the capacity limits of fibers accounting for the presence of fiber Kerr nonlinearity. Some rely on empirical approaches [55]–[58], approximate solutions assuming that fiber nonlinearity is low [59]–[65] or is heuristically considered as a particular form of multiplicative noise [59], [60], [62], while others are limited to specific nonlinear propagation effects [66]. Some fiber capacity studies [67]–[75] solve numerically the equation of propagation in fibers to fully capture all Kerr instantaneous nonlinear effects that include the signal and noise. Some of these studies have been performed using a model with memory [67]–[70] but were limited to modulation formats with at most a few bits per symbol, while others have used modulation formats that are not limited in the number of bits per symbol and are maximally compact [71]–[75]. Capacity limits when using the OFDM format for transmission over fibers has also been reported [76].

In Part 2 of this paper, we discuss in detail the channel model studied here and present the results of fiber capacity estimates using ring constellation that incorporate input constellations optimization, various dispersion maps, and the effect of propagation distance on channel capacity.

VII. HISTORICAL PERSPECTIVE

Experimental demonstrations of high-capacity transmission over a single fiber strand using state-of-the-art technologies have generated considerable excitement over more than two decades. These record capacity experiments are typically reported at the postdeadline sessions of major conferences on optical communication such as the annual Optical Fiber Communications Conference (OFC) or the European Conference on Optical Communication (ECOC). These experiments, often dubbed as “hero” experiments, use the latest technologies in transmitters, receivers, optical fibers, and optical amplifiers to maximize fiber capacity.

The term “capacity,” in the sense of Shannon, is defined as the maximum information rate (averaged over long data sequences) that can be guaranteed for an *arbitrarily* low but nonzero BER. The system is then declared “error free.” For hero experiments, the minimum acceptable BER was initially set to 10^{-9} (at a

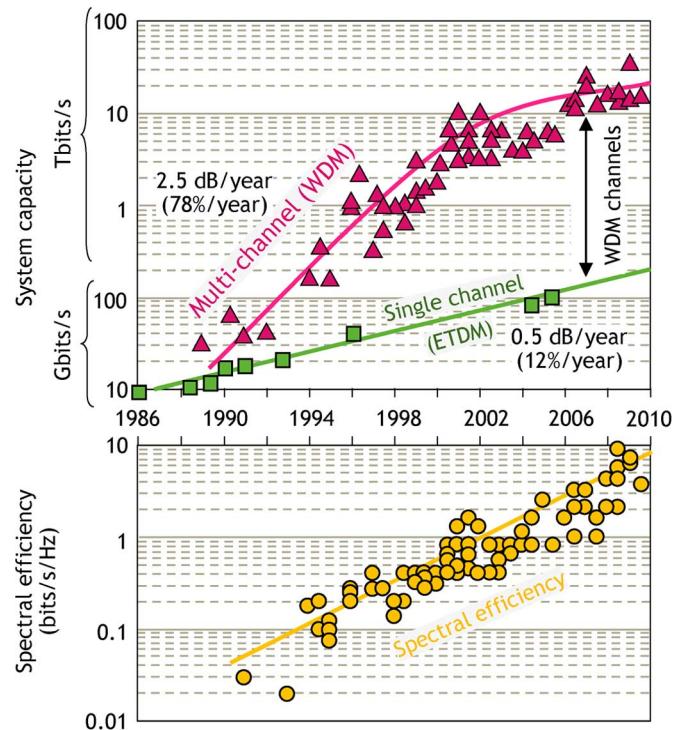


Fig. 20. (Top) Historical evolution of record capacity and (bottom) spectral efficiency of “hero experiments” in fiber-optic communication systems.

time when forward-error correction (FEC) was not used in fiber-optic communication). With the introduction of FEC in the mid-1990s, system experiments started to include a coding bit-rate overhead of typically 7%, and declared “error-free” transmission if the measured BER was at least as good as the value needed at the input of state-of-the-art (hard-decision) FEC devices such that the FEC *output* BER would be between 10^{-12} in the late 1990s to as low as 10^{-16} today. With advances in FEC technology [77]–[79] the required value for the measured *input* BER has shifted from 6×10^{-5} [80] for first-generation 7% Reed–Solomon FEC to 3.8×10^{-3} for second-generation FECs [79]. Codes with higher overheads (on the order of 25%) have been investigated and used mostly in the context of submarine systems so far. They require about a BER of 2×10^{-2} at the FEC input [79]. Note that all these codes guarantee the correction at the prescribed input BER only for uncorrelated errors such as for the AWGN channel. On the nonlinear fiber channel, however, different noise statistics as well as burst errors may be encountered, which can tighten the BER requirements of a code [72], [78], [81]–[84]. This fact is neglected in virtually all “hero” experiments due to the experimental difficulties associated with testing full end-to-end transmission including FEC.

Fig. 20 plots the capacity and the spectral efficiency of hero experiments since the mid-1980s [85]. The lower curve in the top plot shows the transmission bit rate that could be obtained on a single optical wavelength and on a single polarization using electronically time-division multiplexed (ETDM) transmitters. The experienced growth is about one order of magnitude over two decades or about 12% per year, which would have been insufficient to fuel the bandwidth demand of modern data services that grows at about 60% per year [54]. The highest ETDM bit rate reported today is 100 Gb/s

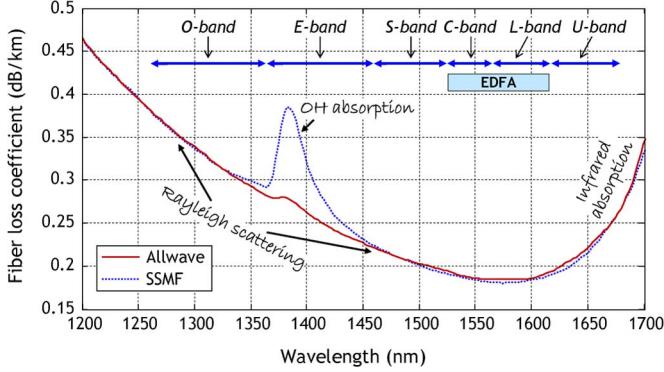


Fig. 21. Spectral dependence of the fiber loss coefficient for a typical low-loss optical fiber (SSMF) and a fiber without the water absorption peak (Allwave). The origin of the main sources of loss are indicated along with the names of the amplification bands and their wavelength ranges. (Courtesy of D. Peckham.)

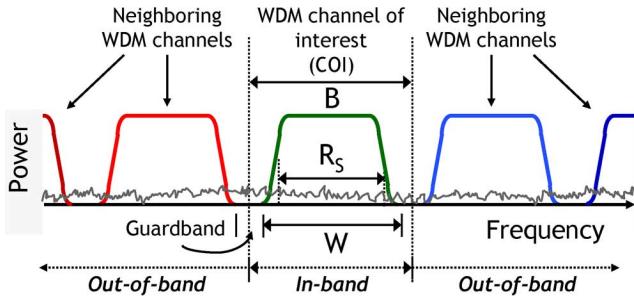


Fig. 22. Spectral layout of the WDM channel superposed to the noisy field originating from ASE.

[86], which is much lower than the bandwidth supported by an optical fiber. The fiber bandwidth that is considered usable for long distance transmission occupies the wavelength range from ~ 1300 to ~ 1700 nm, where fiber loss is moderate or low ($\lesssim 0.35$ dB/km, see Fig. 21). This corresponds to a full channel bandwidth of ~ 54 THz. In practical implementations, the usable bandwidth is limited by the bandwidth of amplification technologies, which is on the order of 5–10 THz (see EDFA box in Fig. 21). Multiple amplification technologies can be used in parallel to form multiband amplifiers and transmission systems [87]–[89] (see the various amplification bands in Fig. 21). The large difference between single-channel ETDM bit rates and the available optical amplification bandwidth suggests dividing the usable fiber bandwidth in smaller frequency bands and populating these bands using WDM [90]. The spectral layout of the WDM channels is shown in Fig. 22 along with the broadband noise generated by optical amplifiers along the path (see Section IX-B). Each WDM channel occupies a bandwidth B that defines the channel spacing. Note that WDM channels are not to be confused with a “channel” in the sense of Shannon and as represented in Fig. 4 of Part 1.

WDM technologies were developed in the mid-1990s and allowed parallel transmission of many WDM channels on the same fiber. The upper curve in the top plot in Fig. 20 shows the total fiber capacity evolution using WDM, with a growth rate of about 78% per year for over 10 years, backed by the steady increase in the bandwidth of optical amplifiers as well as the increase in spectral efficiency due to improvements in laser

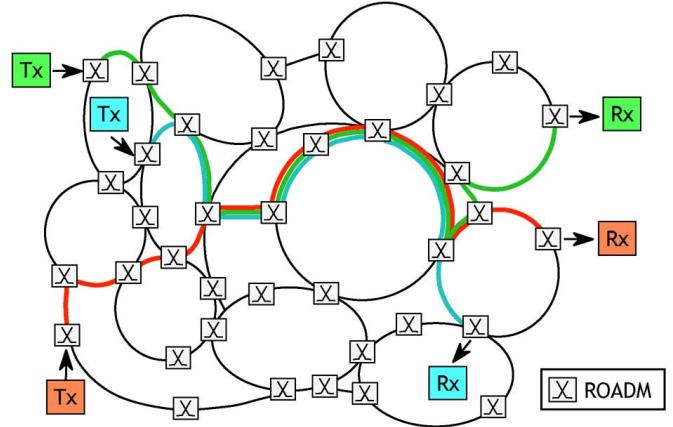


Fig. 23. Schematic representation of an ORN in mesh. The optical functionality of a reconfigurable optical add-drop multiplexer (ROADM) is to multiplex (add) or demultiplex (drop) frequency bands from a WDM signal from one optical fiber to another.

and optical filtering technologies. Around the turn of the millennium, the bandwidth of optical amplifiers approached their maximum values allowed by the material, and the capacity growth began to slow down. Capacity growth became mainly driven by an increase in *spectral efficiency* (see bottom plot of Fig. 20) brought by advanced modulation formats that have quickly been replacing the prevailing OOK systems in the long-haul transport arena.

VIII. OPTICALLY ROUTED NETWORKS

Early fiber-optic transmission systems provided point-to-point transmission [91] with all WDM channels co-propagating over the same optical path. These WDM systems have evolved to now use reconfigurable optical add-drop multiplexers (ROADMs) [92] at network nodes to form *optically-routed networks* (ORNs) [93]–[96] such as represented in Fig. 23.

The granularity of WDM channels has an important effect on the design of ORNs. Because the routing granularity of ROADM cannot be smaller than the granularity of WDM channels unless expensive optical-electronic-optical (OEO) conversion is performed, it becomes economical to establish a hierarchy of nodes in a network. Nodes with insufficient traffic to fill a WDM channel are aggregated in a single, bigger node called a core node. These core nodes are linked together to form a core ORN as represented in Fig. 23. ROADM can then add and drop channels at the WDM channel granularity.

In this paper, we consider an ORN with a generic mesh network topology [94], [95] as schematically represented in Fig. 23. The figure illustrates that independent WDM channels can share some optical fiber spans on their propagation path from their respective transmitters (Tx) to their respective receivers (Rx), distorting each others’ waveforms through fiber Kerr nonlinearity when sharing the same fiber. Individual users generally do not have access to other WDM channels at either the transmitter or the receiver, since this would (in the most general case) involve the exchange of the full optical field information between all transponders physically separated by large distances in the ORN. We further assume that the

total usable fiber bandwidth is entirely filled with uncorrelated WDM channels (see Fig. 30 for a schematic representation of interacting WDM channels), which is what we believe to be the worst-case scenario for a WDM system.

The capacity estimation in this paper is based on the assumption that the usable fiber bandwidth is divided in frequency bands as shown in Fig. 22. We study a typical WDM channel, or channel of interest (COI) from which one can calculate the spectral efficiency (SE), defined in (28), by dividing the capacity of that channel by the bandwidth allocated to it. If desired, the aggregate fiber capacity can then be obtained by multiplying the spectral efficiency by the total fiber bandwidth assumed to be supported by the system. In this study, we will use the term *in-band* to refer to any field that falls within the frequency band of bandwidth B of the COI. Fields outside this frequency band are referred to as *out-of-band*. The guardband represents the difference between the bandwidth B allocated to each channel and the channels' spectral support W at the transmitter. Note that in the nonlinear regime of transmission we are considering (pseudolinear transmission regime [97], [98]), the spectral support is approximatively conserved after transmission (see discussion in Section X-B). Note also that we are considering only unidirectional transmission over one and the same optical fiber strand. Bidirectional transmission is in principle possible and has been reported, e.g., in [99]–[103]. Severe limitations from Rayleigh scattering have been observed in these systems [99], [101].

IX. PHYSICAL MODELS

In this section, we discuss the various physical phenomena present in the fiber channel, including the most important optical and opto-electronic *noise sources*, *Kerr nonlinearity*, and the presence of optical filtering originating from fiber *chromatic dispersion* (CD) as well as from *bandpass filters* (at ROADM)s within ORNs.

A. Signal and Noise in Coherent Optical Receivers

In all studies presented in this paper, we consider ideal *coherent demodulation*, and in particular a perfectly balanced, ideal homodyne receiver. As discussed in Appendix B, ideal homodyne demodulation linearly translates the optical field of the WDM COI into the electronic complex baseband for further processing using standard communication engineering methods. The electrical signal at the receiver output for in-phase (I) and quadrature (Q) components, and prior to any electronic (matched) filtering, reads (Eqs. (94) and (95) of Appendix B)

$$S_I(t) = 2S_D \Re\{E_s(t)E_{\text{LO}}^* + N(t)E_{\text{LO}}^*\} \quad (47)$$

$$S_Q(t) = 2S_D \Im\{E_s(t)E_{\text{LO}}^* + N(t)E_{\text{LO}}^*\} \quad (48)$$

where S_D is the photodetectors' responsivity (in [A/W]); $E_s(t)$ is the complex envelope of the optical signal field; E_{LO} is the temporally constant complex envelope of the LO field acting in the ideal receiver both as a constant gain multiplier as well as a perfect phase reference; $N(t)$ denotes the complex envelope of any other stochastic optical field that is not actively compensated for within the receiver and is therefore contributing to de-

tection noise. The physical origin of the two most important random optical fields, ASE and double-Rayleigh backscatter (DRB), is discussed in detail in Section IX-B.

1) *Beat Noise*: The first term on the right-hand side of (47) and (48) is the desired signal term, and the second term is the beat term between the LO and the optical noise field. Note that this term is the *only* source of beat noise⁷ in this idealized receiver. As is shown in Appendix B, the beat noise between the signal and the optical noise field as well as the beat noise of the optical noise field with itself are fully suppressed by ideal balanced detection. The linear conversion of signal and noise optical fields into the electrical regime also implies that the *statistics* of the optical noise field is fully preserved. In particular, a circularly symmetric complex Gaussian (ccG) optical noise field (such as ASE, as we shall see in Section IX-B1) remains ccG in the electrical domain. The linear translation of optical signal and noise fields into the electronic domain using the LO as a perfect phase reference is the key differentiating factor between coherent demodulation and direct demodulation [105]–[110] with one [111], [112] or more [113] delay interferometers, where the electrical noise will no longer be Gaussian.

The variance of the beat noise term between the LO and the optical noise field is derived in Appendix B and reads

$$\sigma_{S_I, \text{LO}-\text{N beat}}^2 = \sigma_{S_Q, \text{LO}-\text{N beat}}^2 = 4S_D^2 N_0 P_{\text{LO}} B_e \quad (49)$$

where B_e is the power-equivalent bandwidth of the entire receiver opto-electronics (including the matched filter), N_0 is the power spectral density of the noise field $N(t)$ derived in Section IX-B, and P_{LO} is the optical LO power.

2) *Shot Noise*: While the existence of beat noise requires the presence of a stochastic classical optical field in addition to the signal optical field, *shot noise* is always and fundamentally present in any optical receiver. Shot noise is a direct manifestation of the quantum nature of light [114]–[119] and is perceived as random fluctuations of the detected photocurrent, even if the classical optical field by itself is deterministic.⁸ Shot noise in fiber-optic systems cannot be obtained directly from classical optical field descriptions (classical Maxwell's equations); its understanding requires at a minimum a quantized model for light-matter interactions. For example, the power P of ideal, unmodulated laser radiation is a constant, deterministic quantity, resulting in a constant, deterministic photocurrent in a fully classical picture. However, if the quantization of light–matter interactions is taken into account, the photocurrent produced by such a radiation source will exhibit random fluctuations. If we assume a perfectly integrating detector, the statistics of these fluctuations follow a Poissonian probability distribution. For more general receiver characteristics, the shot noise variance is given by [114], [117] $\sigma_{\text{shot}}^2 = 2eS_D P B_e$, where e is the elementary charge. If the optical power is time-varying (as is necessarily the case at the photodetector of all optical communication receivers), the shot noise process itself becomes time-varying and

⁷In optical communications, the term “beat noise” is used to refer to that part of the noise associated with a photodetected signal that directly originates from the classical interference (‘beating’) of optical fields at the photodetector, where at least one of these optical fields is viewed as a stochastic process [104].

⁸If the optical field itself is random, it can be shown [119] that, quite remarkably, the total noise variance splits additively into a beat noise term and a shot noise term.

hence *nonstationary*. The resulting general expressions for the shot noise variance are discussed, e.g., in [104], [117], [119], [120].

In contrast to the cancellation of all but one beat noise term in an ideal, perfectly balanced homodyne receiver, shot noise is generated by the *total* optical power reaching each photodetector, as given by (88) through (93). Owing to the statistical independence of the interactions between classical light and matter within any two detectors, these shot noise fluctuations add up statistically. Hence, we find that for the shot noise variances associated with the two (I/Q) outputs of a balanced homodyne receiver, we have (see Appendix B)

$$\begin{aligned} \sigma_{S_I, \text{shot}}^2(t) &= \sigma_{S_Q, \text{shot}}^2(t) \\ &= 2e(S_1 + S_2)B_e = 2e(S_3 + S_4)B_e \\ &= 2eS_D[|E_s(t) + N(t)|^2 + |E_{\text{LO}}|^2]B_e. \end{aligned} \quad (50)$$

Although shot noise is identified as a nonstationary noise source, we see from this equation that in a coherent receiver shot noise can be made arbitrarily stationary by increasing the LO power, which eventually lets the $|E_{\text{LO}}|^2 = P_{\text{LO}}$ term dominate and dwarfs the shot noise contributions of signal power and noise power. In the limit of $P_{\text{LO}} \gg P_s, P_n$, we are left with

$$\sigma_{S_I, \text{shot}}^2 = \sigma_{S_Q, \text{shot}}^2 = 2eS_D P_{\text{LO}} B_e. \quad (51)$$

Furthermore, we note that at high LO powers, the probability distribution of shot noise rapidly converges from its Poissonian nature towards a Gaussian.

3) Is Shot Noise or Beat Noise More Important to Fiber Capacity: Having specified beat noise and shot noise variances, we are now in the position to answer the question whether shot noise or beat noise is the “more fundamental” noise source in an ideal homodyne receiver when it comes to evaluating fiber capacity. We can establish a fundamental connection between the shot noise and beat noise variances as

$$\frac{\sigma_{\text{LO-N beat}}^2}{\sigma_{\text{shot}}^2} = \frac{2N_0}{h\nu} \quad (52)$$

where we used $S_D = e/h\nu$ for the responsivity of a detector with perfect unity quantum efficiency; h is Planck’s constant, and ν is the optical frequency (ν is used to represent optical frequencies in this paper). Note that (52) is independent of both B_e and P_{LO} . As we will see in Section IX-B, the noise power spectral density N_0 at the receiver for perfectly (ideal) distributed optical amplification is given by $K_T h\nu\alpha L$ [cf. (56)]. Inserting this expression into (52), we see that shot noise is essentially negligible compared to beat noise whenever $\alpha L \gg 1$ where α is the fiber loss coefficient and L the system length, a condition that is very well satisfied for any reasonable fiber-optic transport system length. We may neglect shot noise in our further studies on fiber-optic transport capacities.⁹

4) Thermal and Electronics Noise: Finally, we acknowledge that practical receivers are fundamentally associated with

⁹In certain other applications, e.g., in optical satellite communication links [18], the optical noise power spectral density at the receiver can be much lower than in an amplified fiber-optic system, which can make shot noise the dominant noise contribution in such systems.

thermal noise as well as with other noise sources of purely electronic origin (transistor shot noise, $1/f$ noise, etc.) [121], [122]. We lump all these statistically independent noise sources into an “electronics noise” term with variance σ_{elec}^2 at the two outputs of the balanced receivers. However, as we have seen above, the variance of both shot noise as well as LO-N beat noise can be made arbitrarily large by choosing high enough LO powers, which in turn dwarfs the electronics noise component. As a consequence, while being of great practical importance in designing opto-electronic receiver circuits, electronics noise is not of fundamental interest for a capacity limit estimate and is hence neglected in our studies.

B. Optical Noise Fields

The most important optical noise fields related to fiber-optic transmission are ASE and DRB. We will show here that the most fundamental source of the two is ASE.

Note that signal distortions involving fiber nonlinearity that are not or cannot be compensated for can also be considered a source of “random optical fields” and hence can be viewed as a source of noise albeit with statistical characteristics and correlation properties that can be very different from ASE or DRB. The intricate randomness associated with these nonlinear interactions is taken into account in our work through Monte Carlo simulations, and the effects are discussed in the fiber nonlinearity Section IX-D.

1) Amplified Spontaneous Emission: As mentioned in the context of Fig. 21, the loss coefficient α_{dB} of optical fibers is ~ 0.2 dB/km in the 1550-nm wavelength region. Transmission over a distance L at such wavelengths experiences $\alpha_{\text{dB}} L$ dB of loss. For 2000 km, the accumulated loss is ~ 400 dB, an incredibly large power attenuation of 10^{40} . Clearly, such an enormous attenuation cannot be bridged at a reasonable transmit power (even when leaving aside the detrimental impact of fiber nonlinearities) using modulation formats with a reasonable spectral efficiency in optical fiber. Therefore optical amplification is required along the optical path if frequent opto-electronic regenerations are to be avoided. On the downside, optical amplifiers produce ASE together with signal amplification. One may therefore understand ASE generation in the fiber channel from the fundamental fact that the optical fiber is a lossy transmission medium.

Spontaneous emission is the result of a spontaneous transition from an excited state to a lower energy state in a physical medium, accompanied by the emission of a photon [119], [123], [124]. At the same time, *stimulated emission* is responsible for amplifying a photon within an optical amplifier. Since stimulated emission itself takes place at random, each signal photon passing through an optical amplifier will experience a random multiplication factor, in addition to being accompanied by a number of randomly multiplied spontaneously emitted photons (see, e.g., [110], [125]–[127]). The resulting quantum-mechanical optical field fluctuations are summarized under the term ASE. Remarkably, in [128] and [129], Gordon showed that ASE can be well represented by a random *classical* optical field that has the statistical properties of additive Gaussian noise. The establishment of this equivalence in turn allows the modeling of ASE as a circularly symmetric ccG noise process. A ccG process

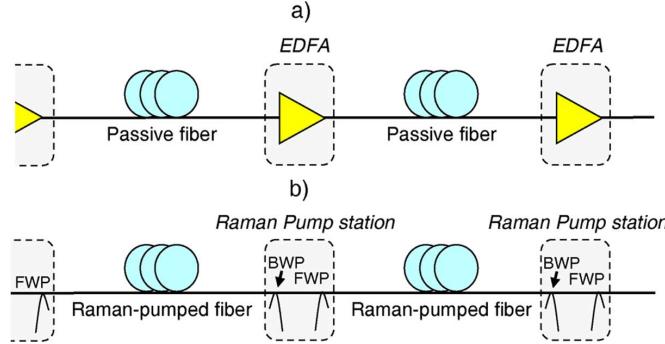


Fig. 24. Two optical amplification schemes for optical transmission over fibers; based (a) on EDFA and (b) on distributed Raman amplification. FWP: forward pumps, BWP: backward pumps.

is fully characterized by its autocorrelation [130]. For ideal distributed Raman amplification, it is given by [124], [131], [132]

$$\mathcal{E}[N(z,t)N^*(z',t')] = N_{\text{ASE}}/L \delta(z-z')\delta(t-t') \quad (53)$$

where $\mathcal{E}[\cdot]$ is the expectation operator [3, Ch. 2], δ is the Dirac functional, and N_{ASE} is the power spectral density of the ASE after a transmission distance L [125], [133], [134] given by (56).

An amplification scheme widely used in fiber-optic communication consists of amplifying the signal periodically at discrete locations along the optical path. This is done by inserting optical amplifiers, generally Erbium-doped fiber amplifiers (EDFAs) [125], to interconnect passive fiber spans. This discrete EDFA amplification scheme is shown in Fig. 24(a). EDFAs are typically unidirectional as they include optical isolators a nonreciprocal component that allows propagation in one direction while blocking propagation in the opposite direction [117], [135]. Fiber span lengths between EDFAs typically range between 40 and 120 km, depending on the network type. This corresponds to between ~ 8 and ~ 24 dB of loss per fiber span before amplification by an EDFA can take place. EDFAs today closely approach the 3-dB theoretical noise figure limit dictated by quantum mechanics.

To improve the OSNR beyond the capabilities of EDFAs, one can transform the passive (lossy) fiber into an amplifying medium by injecting optical pump power (see Fig. 24(b) and Section IX-D2). Such optical pumps provide gain through a stimulated Raman scattering (SRS) process [136], [137] in the transmission fiber and prevent the signal power from dropping along the optical path, which results in improved delivered OSNR [74].

We next calculate the delivered OSNR for both the discrete EDFA and the distributed Raman systems depicted in Fig. 24. For periodically spaced discrete EDFAs, the noise spectral density per state of polarization $N_{\text{ASE}}^{\text{EDFA}}$, generated at the end of a transmission line composed of a chain of N_A amplifiers spaced by fiber spans of length L_A is given by

$$N_{\text{ASE}}^{\text{EDFA}} = N_A(e^{\alpha L_A} - 1)h\nu_s n_{\text{sp}} \quad (54)$$

where α is the fiber loss coefficient given by

$$\alpha = \frac{\alpha_{\text{dB}}}{10 \log_{10}(e)} \quad (55)$$

and $n_{\text{sp}} \leq 1$ is the spontaneous emission factor [125], [133], [134]. The quantity $h\nu_s$ is the photon energy.

For the case of periodically pumped distributed Raman amplification [Fig. 24(b)] we consider *ideal* distributed Raman amplification (IDRA) where the Raman gain *continuously* compensates for the fiber loss [74], [138], [139], i.e., the signal maintains constant average power along the entire transmission span. One can derive the spectral density of noise $N_{\text{ASE}}^{\text{IDRA}}$ of ideal distributed Raman amplification from (54), considering that $N_A = L/L_A$ where L is the length of the transmission line, and by taking the limit $N_A \rightarrow \infty$. We obtain [140]

$$N_{\text{ASE}}^{\text{IDRA}} = \alpha L h\nu_s K_T \quad (56)$$

where n_{sp} is replaced by K_T , the phonon occupancy factor. It is given by $K_T = 1 + \eta(T)$, where $\eta(T) = 1/\{\exp[h(\nu_p - \nu_s)/(k_B T)] - 1\}$ [137] with k_B the Boltzmann constant, T the fiber temperature and ν_p the optical frequency of the Raman pump providing the distributed gain. The factor K_T is approximately 1.13 for Raman amplification of fiber-optic communication systems at room temperature. Experimental demonstrations of nearly ideal distributed gain can be found in [141], [142].

For this capacity limit study, we choose the ideal distributed Raman amplification scheme with Raman gain exactly compensating the fiber intrinsic loss, as it maximizes OSNR (and the SNR) at fixed nonlinear phase which will be discussed in Section IX-D2. The delivered OSNR and SNR can be calculated using (33) and (34).

2) *Double Rayleigh Scattering*: As described in Section X, Rayleigh scattering can be an important source of fiber loss, but it can also be an important source of noise [119], [143]–[145]. A fraction of the Rayleigh scattering of the forward propagating signal is recaptured into the guided mode of the fiber and propagates in the opposite direction of the signal. A fraction of that back-propagating light is then Rayleigh scattered and recaptured into the guided mode of forward propagation, hence co-propagating as *double Rayleigh backscatter* (DRB) along with the signal. This double-scattering process is distributed over the entire fiber length and creates a continuum of echoes that act as multipath interference (MPI) on the signal [146]. Since we deny the receiver knowledge of the amplitudes and phases of these continuum of echoes, MPI is considered as a fundamental source of noise in this context [146].

The power of the DRB light for a lossy fiber with a power loss coefficient per unit length due to Rayleigh scattering of α_R is given by [146]

$$P_{\text{DRB}} = P_{\text{in}}(\alpha_R S)^2 \left[\frac{e^{2(g-\alpha)L} - 2(g-\alpha)L - 1}{4(g-\alpha)^2} \right] \quad (57)$$

where P_{in} is the signal input power to a fiber of length L and S is the dimensionless backscatter recapture fraction that defines how much of the scattered light is recaptured into the guided fiber mode for a particular optical fiber type [137]. The parameters g and α are the distributed gain and fiber loss coefficients per unit of length, respectively, both assumed to be constant along the fiber.

From (57), for ideal distributed Raman amplification for which $g \rightarrow \alpha$, we obtain

$$P_{\text{DRB}} = P_{\text{in}} \frac{(\alpha_R S)^2 L^2}{2}. \quad (58)$$

Depending on the fiber length L , the power P_{DRB} can represent a significant fraction of the signal power P_{in} and limit the effectiveness of Raman amplification for large gain [98], [137], [147]. It is important to note that the DRB power depends quadratically on fiber length when allowing propagation in both directions in a fiber segment. As a result, double Rayleigh backscattering can be reduced dramatically by inserting optical isolators [117], [135], [148]. Dividing the fiber link of length L into N_{ISO} elements of length L/N_{ISO} , the DRB power at the end of the line can now be expressed as

$$P_{\text{DRB}} = P_{\text{in}} \frac{(\alpha_R S)^2 L^2}{2N_{\text{ISO}}}. \quad (59)$$

Making N_{ISO} sufficiently large can reduce DRB to an arbitrarily low level where its impact on capacity is well below that of the ASE. For this reason, DRB is not considered as fundamental a limitation as ASE for our optical fiber capacity evaluation, and we will ignore DRB in the rest of this paper.

C. Optical Filtering

There are two classes of “optical filters” that enter the problem of establishing a fiber channel capacity: all-pass and bandpass filters. The first class is represented by *chromatic dispersion* (CD), originating from the dispersive nature of optical fibers. The second class is represented by the presence of optical bandpass filters at ROADM s to separate and route individual WDM channels in an ORN. These two classes of filters are very different in nature and impact capacity differently.

1) *Fiber Chromatic Dispersion*: There are two distinct origins to the dispersive nature of single-mode optical fibers: material and waveguide [149]. Optical fibers are made of fused silica, a material that exhibits inherent CD. Standard single-mode fibers (SSMFs) have a waveguide dispersion smaller than the material dispersion with a combined dispersion $\approx 17 \text{ ps}/(\text{nm} \cdot \text{km})$. The CD of fibers can be altered dramatically by designing advanced waveguide structures, with waveguide dispersion largely exceeding material dispersion [150, Chs. 2–4], [151, Ch. 2]. For instance, dispersion-compensating fibers with very negative values of total dispersion [e.g., $\approx -120 \text{ ps}/(\text{nm} \cdot \text{km})$] have been engineered using advanced waveguide structures [152]–[156] to compensate the (generally positive) dispersion of transmission fibers [156], [157].

Note that the variation of dispersion with frequency, such as the dispersion slope [158], is neglected here because a fiber can be engineered to have a nonzero value of dispersion over a broad frequency range [159]. The zero dispersion region is generally to be avoided as the effects of fiber nonlinearity are enhanced dramatically [158], [160], [161].

Independent of the origins of dispersion, the equation describing dispersive propagation in fibers can be written as

$$\frac{\partial E}{\partial z} + \frac{i}{2} \beta_2 \frac{\partial^2 E}{\partial t^2} = 0 \quad (60)$$

where β_2 is the group-velocity dispersion (GVD) parameter. CD and GVD are related by

$$D = -\frac{2\pi c}{\lambda^2} \beta_2. \quad (61)$$

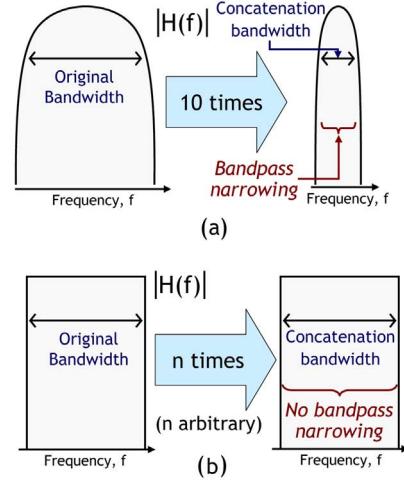


Fig. 25. Concatenation of optical filters. (a) Optical filter with significant amplitude roll-off: spectral narrowing occurs as a result of repeated optical filtering (ten times). (b) Idealized rectangular optical filter: absence of amplitude narrowing for an arbitrary number n of filters.

Equation (60) can be solved in the spectral domain to give,

$$\tilde{E}(z, \omega) = \tilde{E}(0, \omega) \exp(i\beta_2 \omega^2 z/2) \quad (62)$$

where $\tilde{E}(z, \omega) = \int_{-\infty}^{\infty} E(z, t) \exp(i\omega t) dt / \sqrt{2\pi}$ is the Fourier transform of $E(t)$.

As its name suggests, CD produces a spread in time of the various frequency components of a signal due to the difference in group velocity experienced by each frequency component. As CD accumulates, neighboring symbols start to overlap in time, with the number of symbols overlapping increasing with the accumulation of CD. In terms of information theory, CD introduces *memory* to the channel. The memory, expressed in number of symbols, is the spreading in time of a signal of spectral support W and is given approximately by $cWR_s C_D^{\max} / \nu_s^2$ where c is the speed of light and $C_D^{\max} \equiv \max\{|\int_0^z D(z') dz'|\}$ is the maximum excursion of dispersion [162]. This memory expression assumes that the fiber nonlinearity remains sufficiently low.

2) *ROADM Filtering*: Routing individual WDM channels in an ORN (see Fig. 23) requires optical bandpass filters in ROADM s. The number of ROADM s needed to route the signal from a transmitter to a receiver can vary widely in an ORN. In order to accommodate a varying number of ROADM s in the various optical paths, optical filters should be cascadable in their amplitude response. Fig. 25 shows the amplitude response of two types of optical filters. The first type (left plot) has a smooth amplitude roll-off. One can see that concatenating such optical filters (10× in the figure) can result in considerable spectral narrowing. In contrast, an idealized rectangular optical filter can be concatenated an arbitrary number of times n without any spectral narrowing (provided that all filters have the same bandwidth and center frequency). If one chooses “sinc” pulses for modulation (see Fig. 2 and Section II), a rectangular optical filter of bandwidth equal to the symbol rate matches the transmitted signal’s modulation spectrum. An ideal rectangular optical filter suppresses neighboring WDM channels completely.

An ideal rectangular filter is noncausal [25] and hence not physically realizable. However, it is possible to generate filters with characteristics that approach those of an ideal rectangular optical filter as closely as desired by using finite-impulse response (FIR) filters and sufficient time delays to ensure causality [163], [164].

D. Fiber Kerr Nonlinearity

The Kerr effect can be described by a change of refractive index experienced by a medium when traversed by an electric field. It was first observed in 1875 by John Kerr [165] by applying an external magnetic field. In optical fibers, the electromagnetic field of the signal itself can reach a sufficient intensity of 1 GW/m^{20} so as to change the refractive index of the fused silica through the optical Kerr effect. One can distinguish two types of optical Kerr nonlinearity: *instantaneous* and *noninstantaneous*. Both types are discussed in the following.

1) *Instantaneous Kerr Nonlinearity*: The fast change in the fiber medium refractive index that occurs in the presence of an intense electric field is referred to as the *instantaneous* Kerr nonlinearity. Propagation of a signal field $E(z, t)$ in the presence of loss, gain and instantaneous Kerr nonlinearity (neglecting dispersion) can be represented as [98], [158], [166]

$$\frac{\partial E}{\partial z} + \frac{(\alpha - g)}{2}E - i\gamma|E|^2E = 0 \quad (63)$$

where, for simplicity, the z and t dependences of E have been omitted in (63). The nonlinear coefficient γ is given by

$$\gamma = \frac{n_2\omega_s}{cA_{\text{eff}}} \quad (64)$$

where n_2 is the fiber nonlinear refractive index [158], $\omega_s = 2\pi\nu_s$ is the angular optical frequency, c the speed of light and A_{eff} the fiber effective area [158].

Equation (63) has the following *exact* solution

$$E(z, t) = E(0, t) e^{-(\alpha-g)z/2} e^{i\phi_{\text{NL}}(z)} \quad (65)$$

where the integrated nonlinear phase ϕ_{NL} is defined as

$$\phi_{\text{NL}}(z) = \gamma|E(0, t)|^2 L_{\text{eff}}(z) \quad (66)$$

where γ is the nonlinear coefficient and L_{eff} is the effective length defined as

$$L_{\text{eff}}(z) \equiv \frac{1 - e^{(g-\alpha)z}}{\alpha - g} \quad (67)$$

with z represents the fiber length. When gain compensates loss exactly, i.e., $g \rightarrow \alpha$, we have $L_{\text{eff}} \rightarrow z$.

The integrated nonlinear phase for an arbitrary signal power evolution is defined as [158]

$$\phi_{\text{NL}}(z) = \int_0^z \gamma P(z') dz' \quad (68)$$

¹⁰For 100 mW of signal power propagating in a fiber of $100 \mu\text{m}^2$ effective area A_{eff} [158], the intensity of the field $\sim 100 \text{ mW}/100 \mu\text{m}^2 = 1 \text{ GW/m}^2$.

where $P(z) = |E(z)|^2$ is the signal power evolution. Another measure that relates to nonlinear transmission is the integrated nonlinear phase spectral density Φ_{NL} defined as $\Phi_{\text{NL}} \equiv \phi_{\text{NL}}/B$ where B is the WDM channel spacing. The integrated nonlinear phase spectral density is a measure of the nonlinear phase that takes into account the spectral density of the WDM signal.

2) *Noninstantaneous Kerr Nonlinearity*: The noninstantaneous part of the Kerr effect in optical fibers leads to Brillouin [91], [167]–[170] and Raman scatterings [91], [136], [168]–[170]. These processes can be spontaneous [171] or stimulated by the presence of an input wave [145], [172]. Both phenomena can be interpreted as mechanical waves, of low frequencies (acoustic phonons) for the Brillouin scattering and of high frequencies (optical phonons) for Raman scattering. The Raman effect is often modeled as a delayed nonlinear response [173]–[178].

The most important phenomenon associated with stimulated Brillouin scattering (SBS) in fibers is the presence of optical amplification in the backward direction downshifted by about 10 GHz in frequency from the signal and of bandwidth less than 100 MHz [171]. This amplification is generally detrimental to transmission but can be efficiently suppressed using various techniques with minimal impact on capacity, using for instance slow (a few tens of kHz) frequency dithering [179]. It is worth pointing out that SBS gain has some polarization dependence [99], [180]. Stimulated Raman scattering (SRS) also leads to optical amplification, but in contrast to SBS, SRS gain is very broad (~ 10 THz) in fibers. As for SBS, SRS gain is also polarization-dependent [172], [174]. We consider below the two main amplification mechanisms resulting from SRS.

a) *Interchannel Stimulated Raman Scattering*: In WDM transmission, SRS can occur between the different wave-lengths of the WDM spectrum. An important resulting effect of this interchannel SRS is the creation of a tilt of the WDM spectrum [91]. Such a gain tilt results in high-frequency WDM channels to be depleted and the low-frequency WDM channels to be amplified. The WDM spectrum tilt can be calculated using relations derived in [181] and [182]. Some capacity limitations are expected from the WDM spectrum tilt but tilt compensation through gain tilt in the opposite direction and pre-equalization [183] can greatly reduce its impact.

The gain tilt represents only the average effect of inter-channel SRS. The gain provided by interchannel SRS originates from the WDM channels that are data and polarization modulated and experience propagation effects. As a result, the waveform has power and polarization variations in time which produces time-varying gain [170], [184]–[186]. These effects are not incorporated in our analysis. Further studies are needed to assess their importance in the context of fiber capacity.

b) *Ideal Distributed Raman Amplification*: Stimulated Raman scattering can be exploited to create fiber Raman amplifiers by pumping the optical fiber at a frequency about 13 THz above the desired frequency of gain (see [187] and [188]). An important application of SRS in systems is to generate distributed gain [189] in passive transmission fibers. This amplification scheme is referred to as distributed Raman ampli-

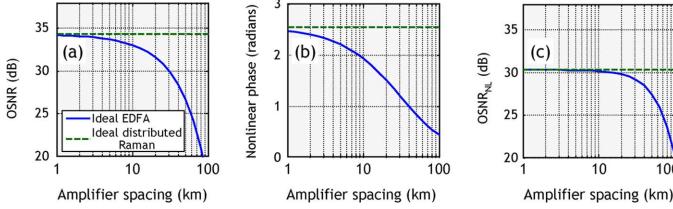


Fig. 26. Delivered OSNR for EDFA and ideal distributed amplification for a typical system: (a) delivered OSNR at fixed input power, (b) nonlinear phase, and (c) delivered OSNR at constant nonlinear phase.

fication, see Fig. 24(b). We will show that such an amplification scheme provides significantly higher delivered OSNR than the discrete Erbium-doped amplification scheme of Fig. 24(a).

We first consider the case with unconstrained signal power and then the case where power is limited to a given value of the integrated nonlinear phase [defined later in (68)] [71]. Without loss of generality, we can initially set the signal power P_{in} to a fixed value of 0 dBm per channel. The amplifier spontaneous emission factor n_{sp} is set to 1 (noise figure ~ 3 dB at large gain [190]) for EDFA and K_T is set to 1 (lowest possible value) for ideal distributed Raman amplification. The system length is 2000 km. Fig. 26(a) shows the delivered OSNR as a function of amplifier spacing (or fiber span length) for fixed launch power for EDFA and ideal distributed Raman amplification. The EDFA system produces lower OSNR even for a span length of a few kilometers. However, a fairer comparison between the two amplification schemes should account for the integrated nonlinear phase of (68) to represent the impact of fiber Kerr nonlinearity. Fig. 26(b) shows the nonlinear phase for the two amplification schemes at a fixed input power and Fig. 26(c) the delivered OSNR when the power P_{in} is adjusted so that transmission takes place at a fixed nonlinear phase.

As seen in Fig. 26(c), for short amplifier spacing (< 20 km), EDFA (full line) and ideal distributed Raman amplification (dashed line) systems produce similar OSNRs at constant nonlinear phase defined as

$$\text{OSNR}_{\text{NL}} \equiv \frac{\text{OSNR}}{\phi_{\text{NL}}}. \quad (69)$$

For large amplifier spacings (50 km and above), the OSNR_{NL} obtained for discrete amplification is much lower than ideal distributed Raman amplification, resulting in a ~ 10 dB difference for 100 km amplifier spacing. This illustrates the benefit of using distributed Raman amplification to achieve low noise at constant nonlinear phase. One can easily show that maximizing OSNR also maximizes SNR using (34).

E. Fiber Propagation

The equation that describes the evolution of the optical field $E(z, t)$ (that contains all WDM channels) in a fiber using ideal distributed Raman amplification (gain continuously compensates fiber loss) with ASE generation can be represented by the

Decomposition of instantaneous fiber Kerr nonlinearities

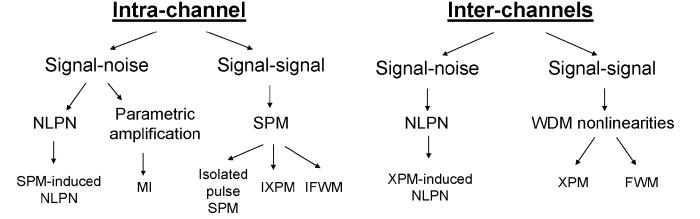


Fig. 27. Decomposition of the instantaneous fiber Kerr nonlinearities into two categories: intrachannel and interchannel nonlinearities. A list of elementary nonlinear interactions for each category is provided. NLPN: nonlinear phase noise, SPM: self-phase modulation, NL: nonlinear, MI: modulation instability, XPM: cross-phase modulation, FWM: four-wave mixing, IXPM: intrachannel XPM, IFWM: interchannel FWM.

stochastic generalized nonlinear Schrödinger equation (GNSE) [131], [132]

$$\frac{\partial E}{\partial z} + \frac{i}{2}\beta_2 \frac{\partial^2 E}{\partial t^2} - i\gamma|E|^2 E = iN(z, t). \quad (70)$$

The terms involving β_2 , γ , and N are the source of fiber channel memory, nonlinearity, and noise generation, respectively. As for (63), for simplicity, the z and t dependences of E have been omitted in (70). The term $N(z, t)$ in (70) is the term describing ASE noise generation for ideal distributed Raman amplification.

The nonlinear interactions resulting from the instantaneous Kerr nonlinearity can be classified in two broad categories, intrachannel and interchannel nonlinearities [97]. The expression intrachannel nonlinearities describes nonlinear interactions involving only fields present in the frequency band of the WDM COI while we refer to interchannel nonlinearities when it involves at least one field outside the frequency band of the COI as shown in Fig. 27.

Each nonlinearity type is further decomposed into signal–signal, signal–noise and noise–noise nonlinear interactions depending on the fields involved in the nonlinear interactions. Further decomposition into elementary nonlinear interactions is possible and is shown in Fig. 27. It is beyond the scope of this paper to discuss in details these elementary nonlinear interactions (see for instance [98], [158], [166] for detailed description). Also, our capacity analysis is general in the sense that they simultaneously include *all* of these interactions, without the need for a decomposition or classification. Nevertheless, such a decomposition may be useful to better understand the specific fields and the specific nonlinear interactions that are responsible for signal distortions, as well as to devise effective means of counteracting them. As mentioned in previous sections, uncompensated signal distortions from the fiber Kerr nonlinearity are considered a source of ‘noise’ that limit capacity.

X. FIBER CHANNEL

This section describes the choices we made in terms of modulation, constellation and digital signal processing to mitigate nonlinear distortions for our fiber capacity estimate.

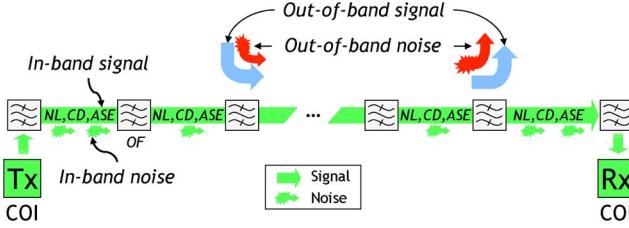


Fig. 28. Schematic of the distribution of physical effects for a COI for the fiber channel considered in this paper. Signal and noise fields are represented for both in-band and out-of-band frequencies.

A. Distributed Impairments

The physical effects present during the propagation from the transmitter (Tx) to the receiver (Rx) are represented schematically in Fig. 28 in the context of where they appear along the propagation path. The signal experiences distributed noise (ASE), fiber nonlinearity, chromatic dispersion (CD), and periodic filtering from optical bandpass filters (OFs). Two out-of-band fields are also shown. They stand as a representation of any other WDM channels added and dropped at random locations in ORNs. In this study, we assume that all neighboring WDM channels co-propagate with the WDM COI during the entire optical path but are not available at either the Tx or Rx. The in-band fields, signal and noise, propagate all the way to the COI receiver while out-of-band fields may be dropped or added along the way. To fully capture and understand the impact of the distributed nature of the impairments of Fig. 28, it requires a full solution of (70).

B. Choice of Modulation

Studies of capacity limits for *band-limited* channels [1] have been developed for linear channels that conserve the signal spectral support. A nonlinear channel can, in general, create new frequencies falling outside the originally transmitted signal spectral support and eventually the channel bandwidth. In the case of the ORN fiber channel, the signal spectrum is repeatedly confined by optical bandpass filters at ROADM. Even though, it is possible to reconstruct a signal truncated by filtering in some scenarios in the absence of noise [191], [192], we surmise that, from a capacity standpoint, it is generally preferable to avoid spectral broadening altogether in a band-limited fiber channel.

Our approach to deal with these difficulties is to place ourselves in a nonlinear regime that limits spectral broadening by fulfilling the condition $L_B(P) \gg L$, where $L_B(P)$ is the signal (power dependent) transmission length over which a non-negligible amount of spectral components are generated beyond the signal spectral support, and L is the transmission length. To make L_B large, we operate in the regime $L_D \ll L_{NL}$, often referred to as the pseudo-linear regime of transmission [97], [98]. The dispersion length is $L_D = T^2/|\beta_2|$ [158] where $T = 1/R_s$ is the symbol duration and β_2 the GVD, related to the dispersion D by (61). The nonlinear length [158] is $L_{NL} = 1/(\gamma P)$ where γ is the nonlinear coefficient defined in (64) and P is the signal power. This regime produces a waveform that changes very rapidly with propagation, helping reduce nonlinear effects. However, the large spreading of the waveform creates a large number of symbols to interact nonlinearly, creating channel memory [162]. One operates in this regime when

transmitting at high symbol rates R_s (~ 25 Gbaud and above) over fibers with high dispersion $|D|$ ($\sim 4\text{ps}/(\text{nm} \cdot \text{km})$ and above) which covers the vast majority of commercial fibers. In our studies, we numerically verified that our capacity estimate results were not limited by spectral broadening for WDM systems as shown later in Fig. 36.

Operating under the conditions of small spectral broadening allows us to use conventional compact spectrum modulation. We studied a range of modulations with square-root raised-cosine spectra with different roll-off factors ρ from 0 to 0.25 (see Section II-B). The primary impact on SE of using a nonzero ρ is a reduction by a factor of $(1 + \rho)$ of the spectral filling to avoid spectrum overlap. We performed SE estimates with low values of ρ of 0.02, 0.01, and 0 and found no significant differences besides the spectral filling factor. For this reason, we use $\rho \sim 0$ or “sinc” pulses described in Section II for essentially all calculations in this paper. Moreover, we assume that the lasers linewidths have negligible impact.

C. Choice of Constellation

Since accurate numerical solutions of (70) are involved and take a long time, it is exceedingly difficult to gather enough numerical statistics for all constellation points individually. As a consequence, we make use of the statistical rotational invariance of the AWGN channel and surmise that this invariance also applies to the nonlinear case, as given by (70). By *statistical rotational invariance* we mean that the channel probability distribution has the property that $p(E_{out}|E_{in}) = p(E_{out}e^{i\Delta\phi}|E_{in}e^{i\Delta\phi})$, i.e., for a certain transmit symbol $E_{in} = re^{i\phi}$, sending the symbol $E_{in} = re^{i\phi+i\Delta\phi}$ will produce the same output density but now rotated by $\Delta\phi$ around the origin of the complex plane. The fact that we are using a statistically rotationally invariant ccG process as noise and statistically rotationally invariant symbol constellations as our signal and our interferers fosters our confidence that we are, in fact, dealing with a statistically rotationally invariant situation. This, in turn, allows us to treat all points on the same ring as statistically equivalent, and we can numerically accumulate statistics by considering only the *relative* transmission induced displacement of a constellation point with respect to its transmit angle by back-rotating each received symbol by its corresponding transmit angle.

The process of back-rotation of constellation points used for numerical evaluation of a capacity estimate is shown in Fig. 29(a,b). Fig. 29(a) shows an original constellation at the transmitter. Each point of the constellation is back-rotated to the positive real axis. In the absence of noise, all symbols belonging to the same ring degenerate to a single point on the real axis. With noise, there is a spreading of the points on each ring that now form “clouds” [see Fig. 29(c)] associated with each ring. With noise and nonlinearity [see Fig. 29(d)], there is a common rotation for all points and further spreading of the clouds. The average rotation of the clouds is referred to as $\bar{\Phi}_{XPM}$. The noisy and nonlinearly distorted clouds after transmission are fitted for each ring to bivariate Gaussian probability distribution functions (PDFs) whose covariance matrices also capture the non-circularity of clouds due to nonlinear signal distortions. From the discretized version of these PDFs, one can calculate capacity estimates using (37). We also explored fitting various

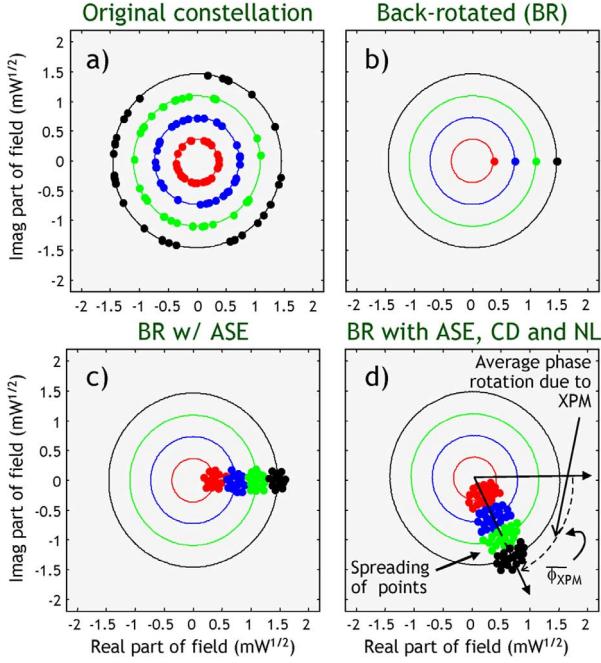


Fig. 29. Example of a four-ring constellation used for numerical evaluation of capacity: (a) original constellation; (b) after back-rotation (BR) without any impairment; (c) after back-rotation with ASE; and (d) after back-rotation with ASE, CD, and fiber nonlinearity.

types of non-Gaussian PDF shapes, both with and without circular symmetry. These PDFs led to very marginally different capacities than obtained for the bivariate Gaussian PDFs. The AWGN channel capacities for ring constellations calculated using the numerical technique described give identical results to Fig. 16 obtained using a semianalytic approach derived in Appendix A.

D. Nonlinearity Compensation Using DSP

Digital signal processing (DSP) can compensate signal distortions in the electrical domain either as pre- or post-equalization (see Fig. 4). Such compensation needs to be performed in the context of capacity calculations for ORNs. Fig. 30 shows a schematic representation of various fields present in the optical path of a WDM COI (labeled 6 in Fig. 30). The ASE noise generated in the same frequency band as the COI, the “in-band noise,” is shown at the bottom of the plot (labeled 7). The signal of the COI is available at both the transmitter (Tx) and receiver (Rx) while the noise being only available at the receiver. The “out-of-band fields” are shown by the first five fields of Fig. 30. Some fields may be available at the transmitter (labeled 5), or at the receiver (labeled 3 and 4) or at neither the transmitter nor receiver (labeled 1 and 2). The scenario that we believe limits the capacity of ORNs the most is when only the in-band fields are available at the transmitter and receiver. This is the ORN scenario we consider in this paper. Because only the in-band fields are assumed available, we focus our attention on compensation of nonlinear distortions from intrachannel nonlinearities. As shown in Fig. 27, one can separate intrachannel nonlinearities into nonlinear interactions that involve only the signal itself (signal-signal intrachannel nonlinearities) and all

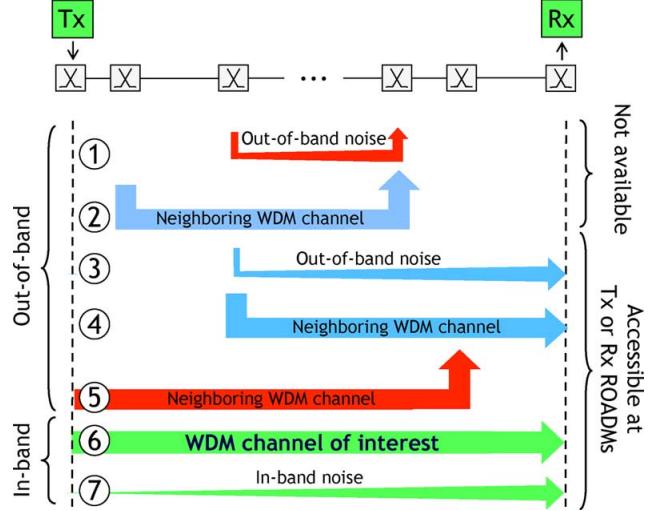


Fig. 30. Schematic representation of interacting fields in nonlinear fiber transmission.

other intrachannel nonlinear interactions that involve the noise (signal-noise and noise-noise).

We compensate intrachannel nonlinear interactions by using reverse propagation (or “back-propagation”) on the fields present in the COI bandwidth. In the absence of optical noise and optical bandpass filtering, back-propagation can undo *exactly* the simultaneous impact of nonlinearity and dispersion on the signal. Back-propagation can be applied at the transmitter, at the receiver, or both. The back-propagation equation can be obtained by setting the right-hand side of (70) to zero and changing z to $-z$. The new equation can be solved in the electronic domain by using digital signal processing to implement the split-step Fourier method (SSFM) [158], which consists essentially of a succession of fast Fourier transforms (FFTs) and complex multiplications (the SSFM is explained in [158, Sec. 2.4.1]). Given a sufficient number of steps of the SSFM, all signal–signal intrachannel nonlinearities can be undone by this process. We will see in Section XI-E that the power levels where signal–noise and noise–noise nonlinear interactions as well as the presence of optical filtering starts to reduce the effectiveness of back-propagation is almost an order of magnitude higher than the power at which WDM nonlinearities limit capacity. Note that we assume that the parameters β_2 , γ , n_{sp} , ν_s , ν_p , and α are known at both the transmitter and receiver.

XI. FIBER CAPACITY RESULTS

In our calculations of fiber capacity estimates for ORNs, we consider SSMF whose parameters are given in Table I. CD of $D = 17 \text{ ps}/(\text{nm} \cdot \text{km})$ corresponds to $\beta_2 = -21.67 \text{ ps}^2/\text{km}$ using the relation given in (61). The amplification scheme is ideal distributed Raman amplification with the characteristics given in Table II. We assume that there is a ROADM at every Raman pumping station (see Fig. 24), where dispersion compensating fibers can optionally be inserted. The signal and optical bandpass filter characteristics in ROADMs, multiplexers, and demultiplexers are given in Table III. We restrict ourselves to a single state of polarization for the signal and most of the numerical simulations were performed with copolarized noise only. It was verified that the same results were obtained when

TABLE I
FIBER PARAMETERS

Dispersion D	17 ps/(nm-km)
Dispersion slope	Set to zero
Loss coefficient α_{dB}	0.2 dB/km
Nonlinear refractive index n_2	$2.5 \times 10^{-20} \text{ m}^2/\text{W}$
Effective area A_{eff}	$80 \mu\text{m}^2$
Nonlinear coefficient γ	$1.27 (\text{W}\cdot\text{km})^{-1}$

TABLE II
AMPLIFICATION PARAMETERS

Amplification scheme	Ideal distributed Raman
Fiber span	100 km
Signal frequency ν_s	193.41 THz ($\lambda_s = 1550 \text{ nm}$)
Spontaneous emission factor n_{sp}	1
Raman Pump frequency ν_p	206.75 THz (1450 nm)
Temperature T	300 Kelvin

TABLE III
SIGNAL AND OPTICAL FILTER PARAMETERS

Number of WDM channels	5
Number of symbols	2048 to 8192
Symbol rate R_s	100 Gbaud
Guardband	2 GHz
(De-)multiplexer bandwidth	102 GHz
ROADMs optical filter bandwidth	102 GHz
Optical filter shape	Rectangular passband
Optical filter locations	Every 100 km

TABLE IV
MODULATION AND CONSTELLATION PARAMETERS

Modulation	'Sinc' pulses
Constellation amplitudes	Equi-spaced rings
Number of rings	1, 2, 4, 8, 16
Constellation angles	Ultra-fine angular grid

polarization effects were included. Finally, the modulation parameters are given in Table IV.

We modeled transmission with a large number of WDM channels and found that increasing the number of WDM channels beyond five only slightly impacts our capacity calculations for the parameters considered; we, therefore, use five WDM channels in the following calculations and study the central channel as our COI. We use constellation points that are randomly chosen on the ring constellation structures, using time sequences varying from 2048 and 8192 symbols per simulation. The large computation time prevented using larger numbers of points but repeated trials with different noise, data realizations and time offsets (including time offsets of a fraction of symbol duration T) led to variations in capacity estimates of only a few tenths of bits/s/Hz. Back-propagation applied at both the transmitter and receiver in variable ratios also produced capacity estimates within a few tenths of bits/s/Hz for all these scenarios presented here.

A. Conventional Dispersion Map

Optical transmission systems that are limited mainly by single-channel nonlinear transmission generally greatly ben-

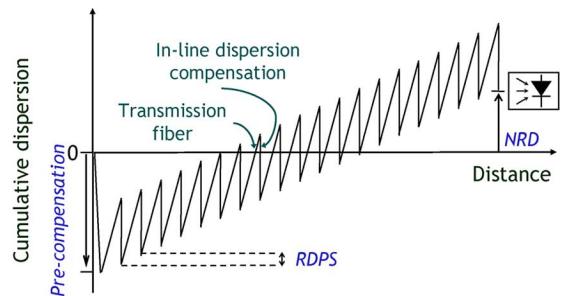


Fig. 31. Definition of parameters in singly periodic dispersion maps. RDPS: residual dispersion per span, NRD: net residual dispersion.

TABLE V
CONVENTIONAL DISPERSION MAP PARAMETERS

Dispersion pre-compensation	-1050 ps/nm
Residual dispersion per span	20 ps/nm
Net residual dispersion at the receiver	0 ps/nm
Nonlinearity in dispersion compensation	None
Spectral dependence of dispersion compensation	Broadband

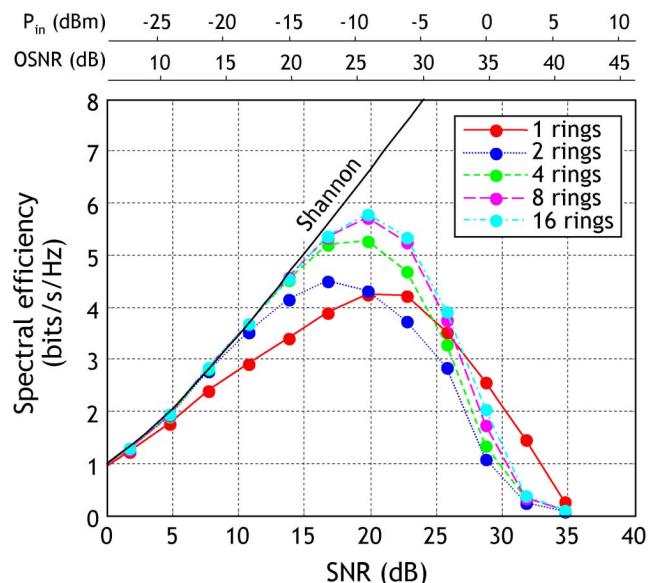


Fig. 32. Spectral efficiency after transmission over 2000 km for uniform ring constellations and a conventional dispersion map.

fit from dispersion mapping where some level of dispersion compensation is periodically applied along the link [98], [193]. The parameters defining a dispersion map are shown in Fig. 31. Optimized values [194] of the parameters of a singly-periodic dispersion map for ideal distributed amplification *in the absence* of nonlinearity compensation are given in Table V. Back-propagation is then used after coherent detection to compensate for CD and fiber nonlinearity.

The fiber channel capacity per unit bandwidth [i.e., the spectral efficiency, SE, defined in (28)] for the system studied here is displayed in Fig. 32 for various numbers of rings [72]. For each ring constellation, the capacity increases following its AWGN capacity (see Fig. 16) at low SNRs ($\lesssim 15$ dB). At moderate SNR (~ 20 dB), the capacities for each ring constellation peak and eventually decrease at higher values of SNRs ($\gtrsim 25$ dB). Note that the signal power level is displayed in Fig. 32 because, unlike for the AWGN channel, not only the SNR but also the ab-

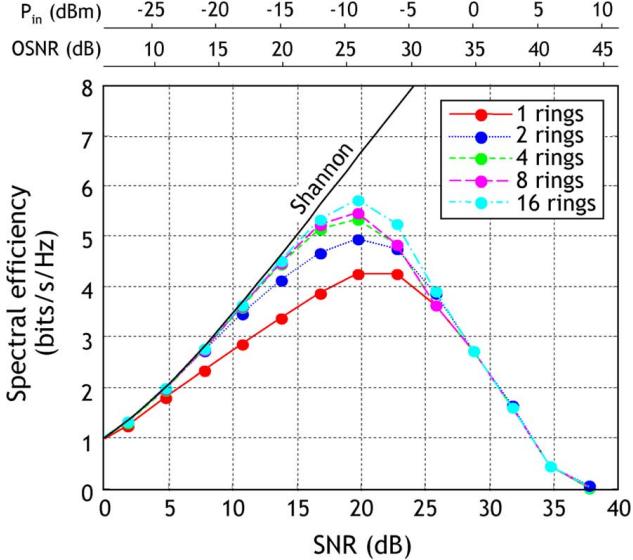


Fig. 33. Spectral efficiency after transmission over 2000 km for ring constellations optimized as described in this paper.

solute signal power is required to calculate fiber capacity. The maximum spectral efficiency of bits/s/Hz is achieved using 16 rings slightly above the capacity for 8 rings. We note that the capacity of the one-ring constellation cuts across the capacities of richer constellations, suggesting that additional capacity may be available in that regime by optimizing the input constellation; this is discussed in the following sections. The statistical variations in capacity near the peak is about 0.2 bits/s/Hz. This was estimated from ten simulations of each case: 1) different noise seeds, 2) different random data, and 3) different timing between WDM channels. The largest variations were observed from different data. Slightly larger variations have been observed beyond the capacity peak, a region of large nonlinear distortions. One should point out that even though we present the fiber capacity results as a function of SNR, unlike for the AWGN channel, the signal power and noise levels may need to be considered separately for the fiber channel.

B. Constellation Shaping

The capacity calculations in Section XI-A use “uniform” ring constellations, i.e., the ring radii are an integer multiple of the inner ring radius, and each transmitted symbol is taken from any of the rings with equal probability (i.e., we assume an equal probability of occupation on each ring). As pointed out in Part I of this paper, the unconstrained channel capacity involves an optimization of the input constellation [see (13)]. We now optimize the ring constellations by varying the ring spacing and the frequency of occupation on each ring. The optimization is summarized in Table VI and a detailed description can be found in [195]. The capacity results with optimized constellations are shown in Fig. 33. Here, the multiring cases always exceed the one-ring capacity values, as expected for optimized constellations (compare to the crossings observed in Fig. 32). At high SNRs, the capacities of the optimized multi-ring constellations become identical to that of the one-ring constellation, as the multiring constellations degenerate to a one-ring constellation for severe nonlinear distortions [195]. The optimization hardly

TABLE VI
CONSTELLATION OPTIMIZATION PARAMETERS (SEE [195])

One ring	Same as uniform ring constellation
Two rings	Optimized both radii and frequency of occupation
> 2 rings	Equal ring spacing with optimized ratio of inner to outer ring radii

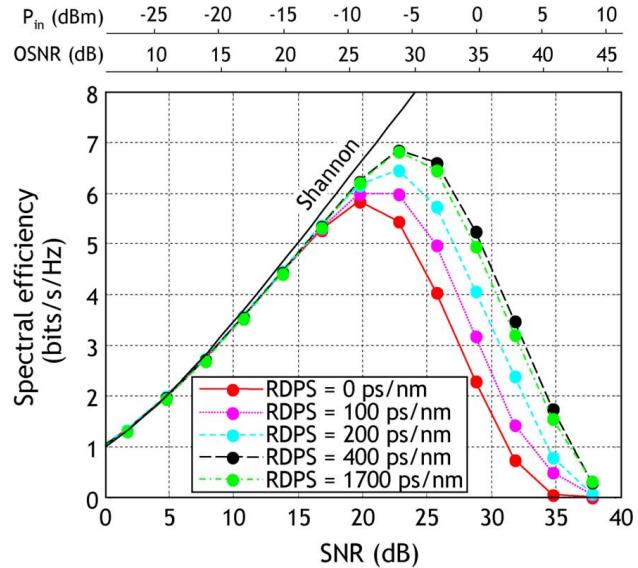


Fig. 34. Spectral efficiency after transmission over 2000 km for various residual dispersion per span.

benefits constellations with more than two rings, so the maximum capacity is not increased. Note that the calculations shown in Fig. 33 were performed with different data realizations (i.e., WDM waveforms) than in the uniform ring constellation case of Fig. 32 resulting in slight statistical variations in capacities for the one-ring case.

C. Effect of Dispersion Map

A measure of the impact of dispersion mapping is presented in Fig. 34 for 16-ring constellations, a number of rings sufficient for our capacity estimate. The residual dispersion per span (RDPS) has been varied from full dispersion compensation per span (RDPS = 0) to the total absence of any in-line dispersion compensation [RDPS = 1700 ps/(nm · km)]. In each case, a dispersion precompensation equal to half the accumulated link dispersion is used and the dispersion is brought back to zero before coherent detection. One observes that increasing the value of RDPS (reducing in-line dispersion compensation) increases capacity, with the maximum capacity being reached in the absence of dispersion compensation.

Two reasons explain this behavior. The first is that back-propagation eliminates all signal–signal intrachannel nonlinearities rendering extraneous the function of the dispersion map to reduce intrachannel nonlinearities in such systems. The second reason is that periodic in-line broadband dispersion compensation re-correlates WDM channels in time producing a “coherent” accumulation of nonlinear distortions rather than a statistical averaging when no realignment occurs in the absence of in-line dispersion compensation. A coherent addition of impairments is more damaging to capacity than a random addition of the same

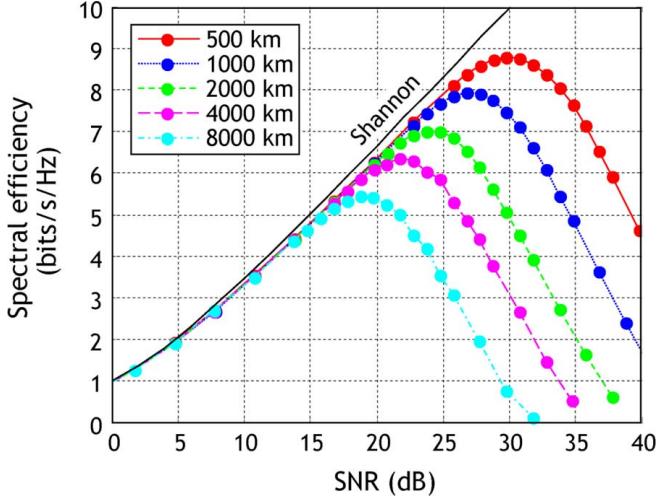


Fig. 35. Spectral efficiency after transmission for various distance. All links are without dispersion compensation.

impairments. It is possible to avoid this recorrelation of WDM channels by using dispersion compensators that are “channelized,” i.e., that compensate dispersion independently for each WDM channel without compensating the relative time delay between them [196]–[199].

The optimization of the input constellation described in Section XI-B and [195] has been applied to systems without in-line dispersion compensation. We were unable to observe any statistically significant increase in capacity.

D. Effect of Distance

We evaluated the dependence of capacity on distance, from 500 to 8000 km. The capacity results are displayed in Fig. 35. The number of rings is 16 for all distances and the number of symbols used here is 8192. We verified by calculating the capacity of 32 rings for a few SNR points that the number of rings was sufficient to capture the maximum capacity, even for 500 km. One observes that the SNR at which the capacity peaks decreases by 3 dB for every doubling in distance. Since the noise level also increases by 3 dB when doubling the distance, the optimum signal launch power is virtually independent of distance. This can be understood by realizing that higher capacities are achieved using richer (denser) constellations that are more sensitive to nonlinear distortions, preventing raising the signal power even when transmission distances are shortened.

E. Origin of Capacity Limitations

To determine the origin of fiber capacity limitations, we calculated the capacity of various signal and noise scenarios for the 2000 km case (the middle curve of Fig. 35). The scenario (1) in Fig. 36 is exactly the middle curve of Fig. 35, which corresponds to the case of WDM transmission with noise (ASE) and optical filters (OFs) every 100 km, and no in-line dispersion compensation. Scenario (2) is an *unphysical case* where ASE is neglected. At high signal power (SNR greater than 30 dB), the capacity is identical to scenario (1) where ASE is included. This indicates that capacity is limited by signal–signal interchannel nonlinearities (see Fig. 27) at these SNR values. At lower SNRs, the capacity increases, even above the linear Shannon limit, since

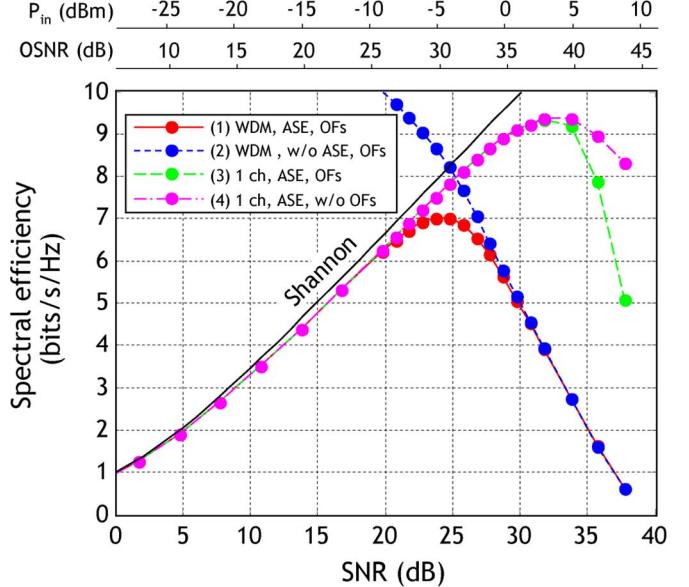


Fig. 36. Spectral efficiency for four signal and noise scenarios for the 2000 km transmission of Fig. 35. (ch: channel.)

there is no ASE to limit capacity for this unphysical scenario. The last two scenarios are for single-channel transmission that both have higher capacities than the first scenario.

The capacity of the single-channel case with OFs [scenario (3)] rolls-off more abruptly than if the OFs are removed [scenario (4)] indicating that the spectral truncation from the presence of optical bandpass filters limits single-channel transmission more strongly than nonlinear distortions between signal and noise (see Fig. 27). Note that single-channel transmission can operate at around 10 dB higher SNR (i.e., power) than the WDM cases. Probably, the most important understanding garnered from Fig. 36 is that signal–signal interchannel nonlinearities are responsible for limiting capacity in the WDM systems considered, and that a moderate increase in capacity can be gained if WDM nonlinear effects could be suppressed.

It is interesting to study the output constellations after transmission for various powers around the optimum capacity. All constellation points are individually back-rotated by their transmit angle since we assume rotational symmetry (see Section X-C). These constellations are displayed in Fig. 37. We show four-ring constellations here to facilitate visualization and because the shape of the clouds are similar to the 16-ring case that produces slightly larger capacities. At low powers ($P_{in} = -15$ dBm per channel), the SNR is low, explaining the large sizes of the clouds. The cloud sizes decrease with increasing signal power until at high powers ($P_{in} = 0$ dBm) the clouds increase in size due to nonlinear effects. The capacity has already decreased when this power level has been reached.

F. Comparison to Record Capacities

The spectral efficiency of recent record experiments that have propagated over more than 300 km and operated at 50 Gb/s and above are shown in Fig. 38. The capacity limit estimate for 500 km transmission over SSMF with a loss coefficient of $\alpha_{dB} = 0.2$ dB/km and effective area of $80 \mu m^2$ is shown for comparison. The record experiments are about a factor three

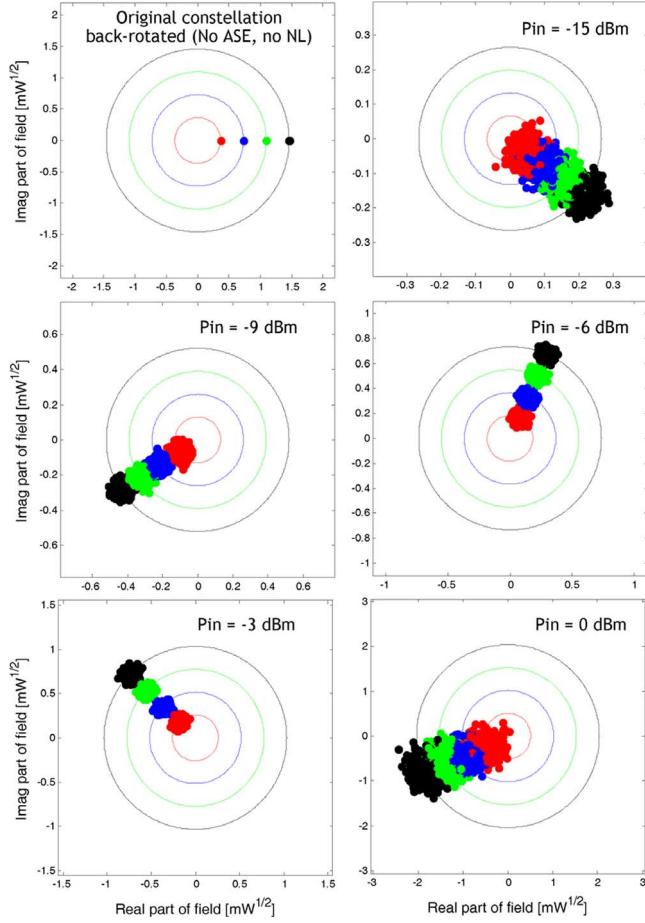


Fig. 37. Four-ring constellations (back-rotated) in the absence of ASE and nonlinearity, and after WDM transmission for various values of input power per channel P_{in} . The optimum power for our capacity estimate is between ~ -6 and $\sim -3 \text{ dB} \cdot \text{m}$.

from the calculated capacity limit estimates. Note however, that record experiments most often use state-of-the-art optical fibers that have larger effective areas than $80 \mu\text{m}^2$ and loss coefficients lower than 0.2 dB/km .

XII. DISCUSSION

We refer to the evaluation of a fiber capacity in this paper as an estimate because of approximations made along the way. First and foremost, the *general* problem of estimating the capacity of a nonlinear channel that includes distributed Kerr-type nonlinearity, dispersion, noise and optical bandpass filtering has had no general formulation yet. In this paper, we operate in a propagation regime (pseudolinear transmission [97], [98]) where the signal spectrum remains highly confined. We use this property to justify using information theory for band-limited AWGN channels. We are also not searching *systematically* for an optimum input distribution, potentially overlooking some capacity gain. In addition, our capacity calculations do not attempt to take advantage of any memory remaining in the channel after back-propagation is used, potentially underestimating capacity.

For instance, one possible way to better approach capacity for a channel that has memory, with or without back-propagation,

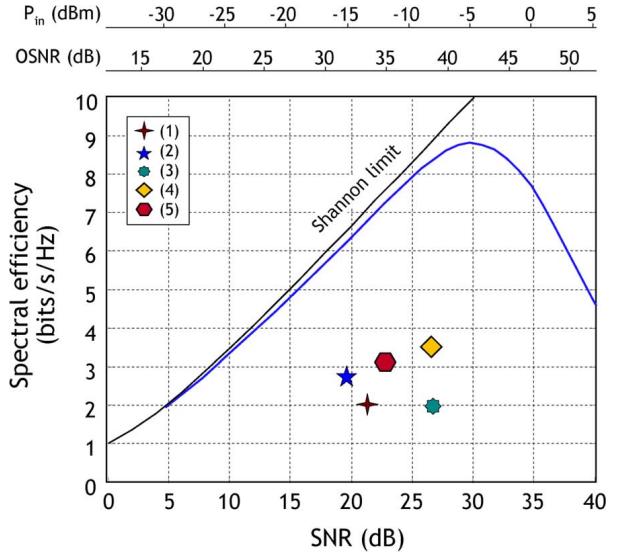


Fig. 38. Spectral efficiency results for recent record experiments. The capacity limit estimate curve for 500 km transmission of Fig. 35 is shown for comparison. There is about a factor three between the capacity limit estimate and the record capacities. The experimental data, labeled (1)–(5) in the legend, are from [200]–[204]. The upper axes apply only to the capacity limit estimation curve.

is by constituting K different subchannels $\{P_{Y_k|X_k}\}_{k=1}^K$ by K periodic extractions of symbols in time starting at a different symbol, from the 0th to the $(K-1)$ st. So, for sufficiently large L , subchannel one, $P_{Y_1|X_1}$, involves symbols transmitted in time slots numbered $0, K, 2K, \dots, (L-1)K$. $P_{Y_2|X_2}$ consists of symbols transmitted in time slots numbered $1, K+1, 2K+1, \dots, (L-1)K+1$. $P_{Y_{K-1}|X_{K-1}}$ consists of symbols $K-1, 2K-1, \dots, LK-1$. Any one of these sub-channels, for sufficiently large K , can by itself be treated as an independent memoryless channel, even though the received symbols suffer impairment through memory effects from symbols sent in neighboring time slots. The symbols in $P_{Y_1|X_1}$ can be decoded and any of the now-known nonlinear influence from symbols exclusively in $P_{Y_1|X_1}$ subtracted from the received signals for sub-channels $P_{Y_2|X_2}, P_{Y_3|X_3}, \dots, P_{Y_{K-1}|X_{K-1}}$. Then the symbols for $P_{Y_2|X_2}$ are decoded and the now-known nonlinear influence of the transmitted symbols decoded in channels described by $P_{Y_2|X_2}$ and $P_{Y_1|X_1}$ together subtracted from the received symbols for $P_{Y_3|X_3}, P_{Y_4|X_4}, \dots, P_{Y_{K-1}|X_{K-1}}$, and so on. Finally, $P_{Y_{K-1}|X_{K-1}}$ would have the now-known nonlinear influences from all the decoded subchannel symbols involving $P_{Y_1|X_1}, P_{Y_2|X_2}, \dots, P_{Y_{K-2}|X_{K-2}}$ removed. Since more and more nonlinear impairment is removed, the capacities of the sub channels increase with increasing subscript. By processing in this way it is clear we have increased the capacity over just treating each subchannel without subtracting away impairments due to already decoded subchannels. What we do in this paper is pessimistically use K times the capacity of $P_{Y_1|X_1}$. Pursuing such advanced, but much more involved forms of processing is beyond the scope of this paper. Indeed, as nonlinear interferences are progressively removed, it suggests that progressively more refined signal constellations be used for channels involving $P_{Y_k|X_k}$ with larger indexes k .

One should note that additional capacity is also expected to be available if WDM channels other than the WDM COI are available at the transmitter or receiver for digital signal processing for nonlinearity compensation. The availability of additional capacity is suggested by Fig. 36 that shows single channel transmission having significantly larger capacity limit estimates than WDM systems in ORNs. As mentioned earlier, WDM channels availability is not guaranteed in ORNs (see Fig. 30). In the case of point-to-point transmission, the full WDM spectrum would be available for processing.

Finally, unlike many channels, the fiber channel can be “molded” to increase the capacity of the channel. This can be done for instance by reducing fiber loss or increasing fiber effective area, as discussed in [75].

XIII. CONCLUSION

A framework for the study of the capacity limits of the fiber channel in optically routed networks has been described. Using a series of advanced technologies, including advanced modulation formats, digital signal processing for fiber back-propagation, flat square bandpass optical filters and optimum coding, we showed that a spectral efficiency per polarization of about 9 bits/s/Hz is achievable over 500 km of standard single-mode fibers.

APPENDIX A INFORMATION RATES OF RING CONSTELLATIONS ON AWGN CHANNELS

Consider one ring with $X = \sqrt{P}e^{i\Phi}$, where Φ is uniformly distributed over the interval $[0, 2\pi]$. Using (20), we need to compute the entropy $H(Y) = \mathcal{E}[-\log_2 p_Y(Y)]$. We have

$$p_Y(y) = \int p_X(x)p_{Y|X}(y|x)dx \quad (71)$$

where $y = y_R + iy_I$, $x = x_R + ix_I$, $p_{Y|X}(y|x)$ is given by (16), and we have

$$p_X(x) = \frac{\delta(P - x_R^2 - x_I^2)}{2\pi\sqrt{P}} \quad (72)$$

where $\delta(\cdot)$ is the Dirac- δ functional defined by the equation

$$\int_I \delta(x)dx = \begin{cases} 1, & \text{if } 0 \in I \\ 0, & \text{else} \end{cases} \quad (73)$$

for the open interval I .

We continue by making the change of variables

$$\begin{aligned} x_R &= x_A \cos \phi, x_I = x_A \sin \phi \\ y_R &= y_A \cos \theta, y_I = y_A \sin \theta. \end{aligned} \quad (74)$$

The Jacobian of the (x_R, x_I) to (x_A, ϕ) transformation is x_A , so the integral in (71) becomes

$$\begin{aligned} p_Y(y) &= \int_0^{2\pi} \int_0^\infty p_X(x)p_{Y|X}(y|x)x_A dx_A d\phi \\ &= \frac{1}{\pi\sigma_N^2} e^{-(y_A^2+P)/\sigma_N^2} \\ &\quad \times \int_0^{2\pi} \frac{1}{2\pi} e^{-2y_A\sqrt{P}\cos(\phi-\theta)/\sigma_N^2} d\phi \\ &= \frac{1}{\pi\sigma_N^2} e^{-(y_A^2+P)/\sigma_N^2} I_0\left(\frac{2y_A\sqrt{P}}{\sigma_N^2}\right) \end{aligned} \quad (75)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero. We thus have

$$H(Y) = \mathcal{E}\left[-\log_2\left(\frac{e^{-(Y_A^2+P)/\sigma_N^2}}{\pi\sigma_N^2} I_0\left(\frac{2Y_A\sqrt{P}}{\sigma_N^2}\right)\right)\right]. \quad (76)$$

We compute (76) numerically as follows. We generate a large number of $X = \sqrt{P}e^{i\phi}$ where ϕ is chosen uniformly over $[0, 2\pi]$, and then generate Y by adding complex Gaussian noise. Finally, we compute $Y_A = |Y|$ and the average in (76). For large entries of $I_0(\cdot)$, we use the approximation [39, p. 47]

$$I_0(x) \approx \frac{e^x}{\sqrt{2\pi x}} \quad (77)$$

so that the term inside the expectation in (76) becomes

$$\frac{(Y_A - \sqrt{P})^2}{\sigma_N^2} \log_2(e) + \log_2\left(2\pi\sqrt{\pi Y_A \sqrt{P} \sigma_N^2}\right). \quad (78)$$

A. Information Rates for Several Rings

Suppose next that we have r rings, i.e., we have

$$p_X(x) = \sum_{k=1}^r P_{X_A}(k) \frac{\delta(P_k - x_A^2)}{2\pi\sqrt{P_k}} \quad (79)$$

where $P_{X_A}(k)$ is the probability of choosing ring k that has power P_k . It is easy to check that $H(Y)$ in (76) is now

$$\mathcal{E}\left[-\log_2\left(\sum_{k=1}^r P_{X_A}(k) \frac{e^{-(Y_A^2+P_k)/\sigma_N^2}}{\pi\sigma_N^2} I_0\left(\frac{2Y_A\sqrt{P_k}}{\sigma_N^2}\right)\right)\right]. \quad (80)$$

We compute (80) numerically as described above for one ring.

It remains to maximize $H(Y)$ over $P_{X_A}(\cdot)$ and P_k subject to the power constraint

$$\sum_{k=1}^r P_{X_A}(k)P_k \leq P. \quad (81)$$

Note that $H(Y)$ is concave in $p_X(\cdot)$ which is linear in $P_{X_A}(\cdot)$ as seen in (79). Thus, once the P_k are chosen one could optimize $P_{X_A}(\cdot)$ by using convex optimization methods. The best $P_{X_A}(\cdot)$ will be a function of the SNR.

A simple approach for choosing the P_k is to use uniform spacing in the field, i.e., choose

$$\sqrt{P_k} = k\sqrt{\gamma P} \quad (82)$$

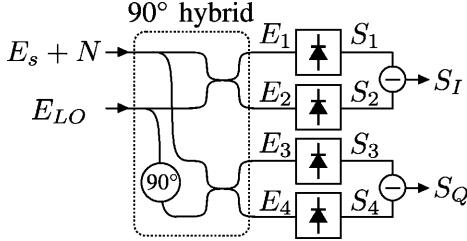


Fig. 39. General setup of a coherent optical receiver. Time variables for the various signals are omitted for visual clarity.

for $k = 1, 2, \dots, r$, where γ is chosen to satisfy (81). For example, if $P_{X_A}(k) = 1/r$ for all k we compute $\gamma^{-1} = (r+1)(2r+1)/6$.

A better approach for choosing the P_k for a large number of rings is to use a spacing that approximates Gaussian signaling in the complex plane. For example, one such choice is a square-root logarithmic spacing with

$$\sqrt{P_k} = \sqrt{\log(2r/(2(r-k)+1))} \sqrt{\gamma P} \quad (83)$$

for $k = 1, 2, \dots, r$, where γ is again chosen to satisfy (81).

APPENDIX B

SIGNAL AND NOISE IN A COHERENT OPTICAL RECEIVER

The structure of a (single-polarization) balanced coherent optical I/Q receiver is shown in Fig. 39. It consists of a 90° optical hybrid, which combines the incident optical field with an LO in both quadratures using two beam splitters as well as two pairs of balanced photodetectors. Their difference signal constitutes the output signal of the receiver.

With reference to Fig. 39, the optical fields at the four detectors can be written in the following form [117]

$$E_1(t) = \sqrt{1-\varepsilon}[E_s(t) + N(t)] + \sqrt{\varepsilon}E_{\text{LO}} \quad (84)$$

$$E_2(t) = -\sqrt{\varepsilon}[E_s(t) + N(t)] + \sqrt{1-\varepsilon}E_{\text{LO}} \quad (85)$$

$$E_3(t) = \sqrt{1-\varepsilon}[E_s(t) + N(t)] + i\sqrt{\varepsilon}E_{\text{LO}} \quad (86)$$

$$E_4(t) = -\sqrt{\varepsilon}[E_s(t) + N(t)] + i\sqrt{1-\varepsilon}E_{\text{LO}} \quad (87)$$

where the sign of the signal term originates from energy conservation within the lossless beam splitters with power transmission ε (ideally, $\varepsilon = 0.5$), and the multiplication by i is due to the 90° phase shift of the LO within the 90-degree optical hybrid. After square-law photodetection with responsivities $S_{D,k}$ (in [A/W]; ideally $S_{D,k} = S_D$ for all k), the four electrical signals $S_k(t) = S_{D,k}|E_k(t)|^2$ read

$$S_1 = S_{D,1}[(1-\varepsilon)[|E_s|^2 + |N|^2 + 2\Re\{E_s N^*\}] + \varepsilon|E_{\text{LO}}|^2 + 2\sqrt{\varepsilon(1-\varepsilon)}\Re\{E_s E_{\text{LO}}^* + N E_{\text{LO}}\}] \quad (88)$$

$$S_2 = S_{D,2}[\varepsilon[|E_s|^2 + |N|^2 + 2\Re\{E_s N^*\}] + (1-\varepsilon)|E_{\text{LO}}|^2 - 2\sqrt{\varepsilon(1-\varepsilon)}\Re\{E_s E_{\text{LO}}^* + N E_{\text{LO}}\}] \quad (89)$$

$$S_3 = S_{D,3}[(1-\varepsilon)[|E_s|^2 + |N|^2 + 2\Re\{E_s N^*\}] + \varepsilon|E_{\text{LO}}|^2 + 2\sqrt{\varepsilon(1-\varepsilon)}\Im\{E_s E_{\text{LO}}^* + N E_{\text{LO}}\}] \quad (90)$$

$$S_4 = S_{D,4}[\varepsilon[|E_s|^2 + |N|^2 + 2\Re\{E_s N^*\}] + (1-\varepsilon)|E_{\text{LO}}|^2 - 2\sqrt{\varepsilon(1-\varepsilon)}\Im\{E_s E_{\text{LO}}^* + N E_{\text{LO}}\}] \quad (91)$$

where we omit the time variable t for notational simplicity, the asterisk denotes complex conjugation, and \Re and \Im stand for the real and imaginary parts, respectively. The difference signals at the two outputs of the balanced receivers then read

$$S_I = [S_{D,1}(1-\varepsilon) - S_{D,2}\varepsilon][|E_s|^2 + |N|^2 + 2\Re\{E_s N^*\}] + [S_{D,1}\varepsilon - S_{D,2}(1-\varepsilon)][|E_{\text{LO}}|^2 + 2[S_{D,1} + S_{D,2}]\sqrt{\varepsilon(1-\varepsilon)}\Re\{E_s E_{\text{LO}}^* + N E_{\text{LO}}\}] \quad (92)$$

$$S_Q = [S_{D,3}(1-\varepsilon) - S_{D,4}\varepsilon][|E_s|^2 + |N|^2 + 2\Re\{E_s N^*\}] + [S_{D,3}\varepsilon - S_{D,4}(1-\varepsilon)][|E_{\text{LO}}|^2 + 2[S_{D,3} + S_{D,4}]\sqrt{\varepsilon(1-\varepsilon)}\Im\{E_s E_{\text{LO}}^* + N E_{\text{LO}}\}]. \quad (93)$$

For an ideally balanced receiver ($\varepsilon = 0.5$ and $S_{D,k} = S_D$) all direct detection terms vanish, and we are left with

$$S_I(t) = 2S_D\Re\{E_s(t)E_{\text{LO}}^* + N(t)E_{\text{LO}}\} \quad (94)$$

$$S_Q(t) = 2S_D\Im\{E_s(t)E_{\text{LO}}^* + N(t)E_{\text{LO}}\}. \quad (95)$$

Finally, these expressions are convolved with the opto-electronic front-end's impulse response $h(t)$, which can, e.g., implement a matched filter.

The first term on the right-hand side of (94) and (95) is the desired signal term, and the second term is the beat term between the LO and the optical noise field, which is the *only* beat noise term of relevance. Note that both the beat term between signal and optical noise as well as the noise–noise beat term are fully eliminated by ideal balanced detection. The linear conversion of signal and noise optical fields into the electrical regime also implies that the statistics of the noise are fully preserved. In particular, a Gaussian optical noise field (ASE) will remain Gaussian in the electrical domain, in contrast to direct detection receivers or imbalanced coherent receivers, which generally exhibit non-Gaussian detection noise statistics [105]–[108].

The variance of the beat noise term between the LO and the optical noise field can be directly calculated from (94) and (95), as outlined, e.g., in [119] and [205]. One starts by taking the expectation of the squared magnitude of the beat-noise term

$$\sigma_{S_I, \text{LO-Nbeat}}^2 = 4S_D^2\mathcal{E}[|\Re\{N(t)E_{\text{LO}}\} * h(t)|^2] \quad (96)$$

assuming that $N(t)$ is a zero-mean stochastic process. (An equivalent procedure can be performed for the beat noise in the quadrature component.) Writing out the convolution in its integral form, expanding the real part into the sum of two complex conjugated terms, expressing the squared magnitude as a multiplication with the complex conjugate, assuming the optical noise field to obey $\mathcal{E}[N(\tau)N(\tau')] = 0$, which holds, e.g., for circularly symmetric ccG noise by the moment theorem of Gaussian random variables, and taking note of the fact that E_{LO}^* is temporally constant, we arrive at

$$\sigma_{S_I, \text{LO-N beat}}^2 = 2S_D^2P_{\text{LO}}\Re\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathcal{E}[N(\tau)N^*(\tau')]h(t-\tau)h(t-\tau')d\tau d\tau'\right\}. \quad (97)$$

Further simplifications can be obtained through certain assumptions on the noise autocorrelation $\mathcal{E}[N(\tau)N^*(\tau')]$ and the detector's impulse response. For example, if we assume the noise to be white with power spectral density N_0 over the opto-electronic detection bandwidth, we have $\mathcal{E}[N(\tau)N^*(\tau')] = N_0\delta(\tau - \tau')$, and consequently

$$\sigma_{S_I, \text{LO-N beat}}^2 = 4S_D^2 P_{\text{LO}} N_0 B_e \quad (98)$$

where we made use of $\int h^2(t)dt = \int |H(f)|^2 df = 2B_e$, B_e being the power equivalent bandwidth of the real-valued impulse response $h(t)$. This result agrees with the commonly used beat noise variance approximation derived in [206].

APPENDIX C CAPACITIES AND ERROR RATIOS OF DISCRETE MODULATION SCHEMES ON AWGN CHANNELS

The capacity-achieving input distribution for the AWGN channel under an average power constraint $\mathcal{E}[|X|^2]$ is circularly complex Gaussian (see (22) and Fig. 14). Any other constellation, discrete or continuous, is sub-optimum for the AWGN channel. As we shall see in Subsection D below, one can approach the AWGN channel capacity (23) with discrete constellations by using a sufficiently large number M of well-placed points. Any remaining SNR gap between the capacity (23) and the information rate (37) of constellations is called the *shaping gain* [207], [208]. For example, for M -QAM with uniform input probabilities, the shaping gain grows to 1.53 dB asymptotically for large M (or large SNR). The gap in SNR per information bit between uncoded and coded transmission at a certain error ratio, e.g., 10^{-16} , is known as the *effective* (or *net*) *coding gain* [208].

Discrete constellations are often compared in terms of their uncoded symbol or bit error ratio (SER/BER) performance and their coded information rates (12) or (37) with hard- or soft-decision receivers, respectively. Different constellations may be optimum at different SNR values; generally, a “globally optimum” constellation cannot be found. The problem of placing M points in the signal space such that the capacity at a given SNR is maximized (or the BER is minimized) is nontrivial and attracted much attention in early days of digital communications [209]–[211]. An overview of the historical development of two-dimensional constellations can be found in [212].

For a given constellation, error ratios and capacities depend on how the noisy samples are processed in the receiver. Upon observing the channel output Y , the receiver might make a decision on the transmitted channel input X . The probability of a symbol error is minimized by the *maximum a posteriori* (MAP) criterion, i.e., by a receiver that, for an observed channel output y , decides on that input symbol $x \in \mathcal{X}$ which maximizes $P_{X|Y}(x|y)$. For equally probable input symbols, this is equivalent to a *maximum likelihood* (ML) decision which maximizes $p_{Y|X}(y|x)$ [24]. A receiver that makes such *hard decisions* maps each point in the received signal space onto a discrete symbol; the set of values that are mapped onto a given symbol form its *decision region*. If a received value is outside the decision region of the transmitted symbol, a *symbol error* occurs. For the AWGN channel, the MAP (or ML) criterion reduces to deciding on the symbol that is closest to the received value,

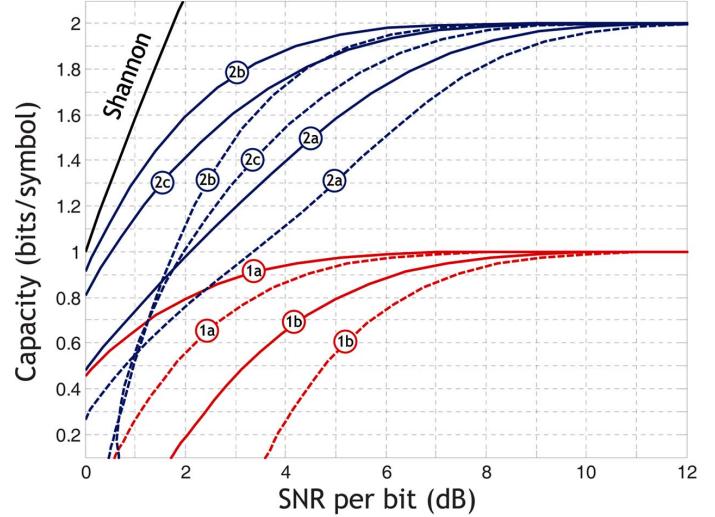


Fig. 40. Capacities of various binary and quaternary formats with soft (solid lines) and hard decision (dashed lines): OOK (1b), BPSK (1a), 2-ASK/2-PSK (2a), QPSK or 4-QAM (2b), ring-1-3 (2c).

i.e., the decision regions are regions of *minimum distance* [24] (so-called *Voronoi* regions [213]).¹¹ The SER of an uncoded hard-decision system is given by

$$\text{SER} = 1 - \sum_{x \in \mathcal{X}} P_X(x) \int_{y \in \mathcal{R}_x} p_{Y|X}(y|x) dy \quad (99)$$

where \mathcal{R}_x denotes the decision region corresponding to input symbol x . For the AWGN channel, $p_{Y|X}(y|x)$ is a complex Gaussian PDF (16).

The most likely error event is to receive y within a decision region that neighbors that of the transmitted symbol. To reduce the BER, the bit sequences assigned to adjacent symbols should differ only in one digit; such an assignment is called a Gray mapping (after Frank Gray, who used the term *reflected binary code* [52]). For constellations with one degree of freedom (ASK, PSK), a Gray mapping is easily found. Square QAM constellations (with even $\log_2(M)$) can be Gray encoded hierarchically by subdividing the constellation into smaller blocks [212]. Similarly, we have produced Gray mappings for ring constellations, i.e., ASK/PSK combinations, by encoding the phase and amplitude separately and concatenating the code words appropriately. There are constellations for which a perfect Gray mapping cannot be found; among them are *cross-QAM* constellations (with odd $\log_2(M)$) [212]. Another simple example for which a perfect Gray mapping does not exist is a quaternary constellation with 3 points on a ring and 1 point in the origin. We call this constellation *ring-1-3* in the following. For a given bit mapping, the BER of an uncoded hard-decision system is calculated by integrating over the decision regions similar to (99) and weighting each probability with the corresponding number of wrongly decided bits (see Fig. 10).

To compare various modulation schemes, it is instructive to express the bit or symbol error ratio, and the capacity, in terms of the *SNR per information bit* or SNR_b as in Fig. 11. Recall

¹¹The minimum distance decision rule is optimum only for circularly symmetric and monotonically decreasing noise PDFs; the noise encountered in optical communication systems, in particular those with direct-detection receivers, may deviate significantly from this “ideal.”

from (31) that $\text{SNR}_b = \text{SNR}/\log_2(M)$ for uncoded BER and SER evaluations.

At the receiver, one can distinguish between *hard decision* and *soft decision* systems (see Section II-C). The latter pass the received continuous channel output y to the decoder. This permits the use of soft-decision error-correcting codes such as Turbo or low-density parity-check (LDPC) codes. In contrast, hard-decision receivers make a definitive decision to which input symbol the output symbol y corresponds to before decoding. The hard decision, which is performed according to the MAP (or ML) criterion as described above, can be regarded as a quantization into Voronoi regions or *Voronoi cells* [213]. In this quantization process, information is inevitably lost, so that the achievable information rate (12) is generally less than the rate achieved by soft decision (37). Note that the bit mapping function of the digital modulator (see Fig. 4) can be an arbitrary bijective mapping without loss of information. Therefore, the achievable information rates for both hard- and soft-decision systems are independent of the bit mapping used by the digital modulator.

A geometric property that is used to characterize modulation schemes is their minimum distance d_{\min} , i.e., the smallest Euclidean distance between any two symbols of the constellation. The values for d_{\min} given in this Appendix are normalized to unity average symbol energy. At large values of SNR, the BER performance is dominated by the minimum distance, whereas this is not necessarily the case for lower SNRs, where a received symbol may be mistaken for a symbol that is further away than the nearest neighbor. A general geometric property of any optimum M -ary modulation scheme is that its center of gravity is the origin, as this minimizes the average symbol energy [22], [210]. Well-known modulation schemes include *amplitude-shift keying* (M -ASK/2-PSK in this paper), *phase-shift keying* (PSK) and *quadrature amplitude modulation* (QAM). A combination of amplitude and phase modulation (ASK/PSK) yields signal constellations where M points are located on r rings, so that there are M/r symbols per ring. Depending on the rings' amplitudes and phase offsets, this scheme describes a large number of constellations, e.g., all different 8-QAM constellations. In earlier references, this scheme is called a *Type I* system [212]. To achieve better minimum-distance properties, it is better to allow a variable number of symbols on each ring (*Type II* in the literature). Such constellations usually have one point in the origin and a growing number of points on the outer rings and are often referred to by the number of points on their rings, e.g., *ring-1-3-5-7* for the 16-ary constellation depicted in Fig. 44 with one point in the origin, three points on the first ring, five points on the second ring, and seven points on the outermost (third) ring. The densest 2-D lattice is hexagonal ("try penny packing" [211]); hexagonal constellations therefore maximize d_{\min} and are asymptotically (i.e., for large SNR) optimal. Such constellations become symmetrical when the outer-most hexagon is completely filled, i.e., for $M = 7, M = 19, M = 37$, etc.

The error ratio and capacity results presented in the remainder of this Appendix have been obtained as follows. To calculate the SER at a given SNR value, the integral in (99) (with $p_{Y|X}(y|x)$ given by (16)) was evaluated numerically for every point $x \in \mathcal{X}$ of the constellation. The decision regions \mathcal{R}_x are regions of minimum distance, i.e., every point in the complex plane that is closer to x than to any other constellation point belongs to

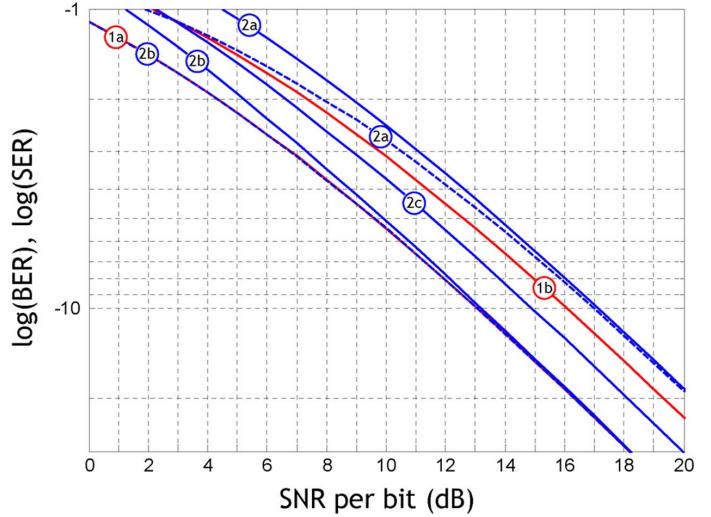


Fig. 41. Symbol and bit error ratios (SER—solid lines, BER—dashed lines) of various binary and quaternary formats: OOK (1b), BPSK (1a), 2-ASK/2-PSK (2a), QPSK or 4-QAM (2b), ring-1-3 (2c). The BER of QPSK (2b) is equal to that of BPSK (1a).

\mathcal{R}_x . Using the same decision regions, the BER is numerically evaluated as

$$\text{BER} = \sum_{x_i \in \mathcal{X}} P_X(x_i) \sum_{x_j \in \mathcal{X}, x_j \neq x_i} P_{ij} n_{ij}, \quad (100)$$

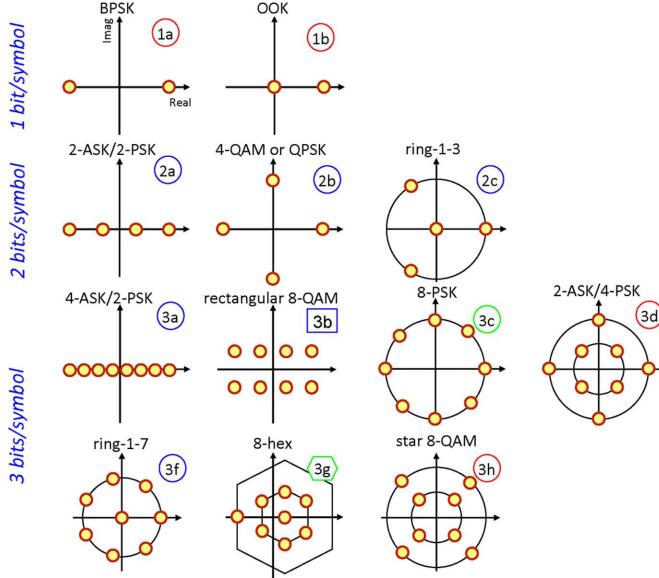
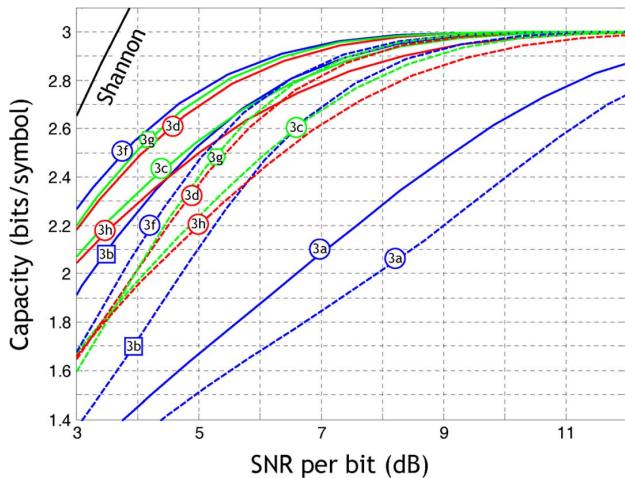
where the probability for receiving x_j given that x_i was sent is given by

$$P_{ij} = \int_{y \in \mathcal{R}_{x_j}} p_{Y|X}(y|x_i) dy, \quad i \neq j, \quad (101)$$

and n_{ij} is the number of bit errors made for that symbol error ($1 \leq n_{ij} \leq \log_2(M)$). Using a Gray mapping decreases this number n_{ij} and hence the BER. To obtain the soft-decision capacity value, (37) was integrated numerically for every input symbol $x \in \mathcal{X}$ and using (16). To obtain the hard-decision capacity, the transitional probabilities (101) are used to evaluate (12).

A. Binary and Quaternary Constellations

For binary ($M = 2$) and quaternary ($M = 4$) signaling, those constellations with the largest minimum distance yield the best performance in terms of error ratio or capacity at all values of SNR. The distance between the symbols for OOK [see Fig. 42(1b)] is $d_{\min} = \sqrt{2}$, whereas BPSK [see Fig. 42(1a)] has $d_{\min} = 2$ (all at unity average symbol energy). QPSK [see Fig. 42(2b)] achieves $d_{\min} = \sqrt{2}$, which exceeds $d_{\min} = 0.894$ for 2-ASK/2-PSK [see Fig. 42(2a)], and $d_{\min} = 1.15$ for ring-1-3 [see Fig. 42(2c)]. Fig. 40 depicts the capacities of these formats for soft- and hard-decision receivers as a function of the SNR per information bit. All schemes achieve $\log_2(M)$ bits/symbol for large SNR. At lower SNRs, QPSK/4-QAM (2b) has the highest capacity, followed by ring-1-3 (2c) and 2-ASK/2-PSK (2a). The dashed lines show the capacity values for hard-decision, so that the difference between corresponding solid and dashed line pairs can be interpreted as the amount of information that is destroyed by the hard decision process

Fig. 42. Examples of constellations with $M = 2$, $M = 4$, and $M = 8$ points.Fig. 43. Capacities of 8-ary formats with soft (solid lines) and hard decision (dashed lines): 4-ASK/2-PSK (3a), rectangular 8-QAM (3b), 8-PSK (3c), 2-ASK/4-PSK (3d) with $r_2/r_1 = 2$, ring-1-7 (3f), hexagonal (3g), and star 8-QAM (3h).

or, equivalently, as the SNR gain that can be achieved by processing soft values in the receiver.

Fig. 41 shows the SER for these formats. For binary constellations, the SER and BER are identical. For $M > 2$, at high SNRs, the dominant error event is the detection of a neighbor symbol. With Gray mapping, this induces only a single bit error, so that for schemes with $M > 2$, BER and SER are asymptotically equal. This can be observed in Fig. 41, where the BER and SER curves of QPSK (2b) and 2-ASK/2-PSK (2a) merge for large SNRs.

B. 8-ary Constellations

Examples for constellations with $M = 8$ points are depicted in Fig. 42.

Using only a single quadrature results in bad performance at $M > 4$; 4-ASK/2-PSK (3a) has a minimum distance of only

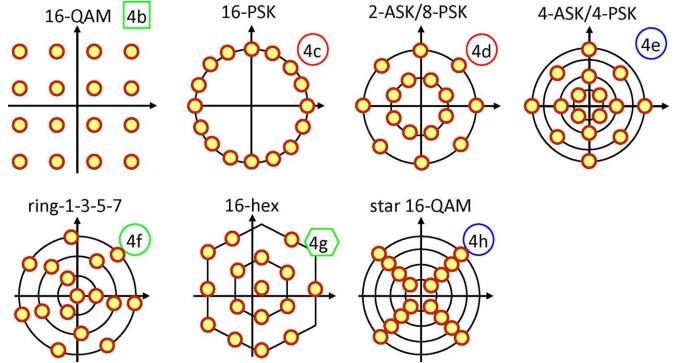
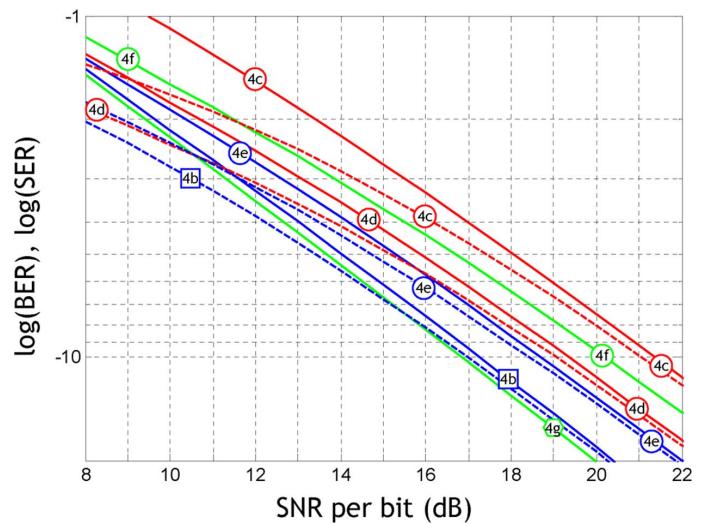
Fig. 44. Constellations with $M = 16$ points.

Fig. 45. Symbol and bit error ratios (SER: solid lines, BER: dashed lines) of 16-ary formats: hexagonal (4g), QAM (4b), PSK (4c), ring-1-3-5-7 (4f), 2-ASK/8-PSK (4d), 4-ASK/4-PSK (4e).

$d_{\min} = 0.436$. This compares to $d_{\min} = 0.765$ for 8-PSK (3c) and $d_{\min} = 0.894$ for 2-ASK/4-PSK (3d) with a ring amplitude ratio of $r_2/r_1 = 2$. 2-ASK/4-PSK (3d) subsumes other well-known constellations. For $r_2/r_1 = \sqrt{2}$, we obtain square 8-QAM with $d_{\min} = 0.816$. Setting $r_2/r_1 = (1 + \sqrt{3})/\sqrt{2} \approx 1.93$ delivers an optimum minimum distance of $d_{\min} = 0.919$. Star 8-QAM (3h) (which is 2-ASK/4-PSK with equal phase angles in both rings) has $d_{\min} = 0.632$. Rectangular 8-QAM (3b), which are eight points on a 2-by-4 grid, achieves the same minimum distance as square 8-QAM, $d_{\min} = 0.816$. More unusual 8-ary formats are ring-1-7 (3f) ($d_{\min} = 0.928$) and 8-hex (3g) ($d_{\min} = 0.963$). The latter is the optimum 8-ary constellation in terms of minimum distance [210]. Notice that the constellation is dc-free; because of its asymmetry, this implies that the innermost constellation point is not located at the origin.

Fig. 43 shows soft- and hard-decision capacity values for 8-ary modulation schemes. It is noteworthy that despite its having the largest minimum distance, 8-hex performs slightly worse than ring-1-7 for a wide SNR range. The SER of 8-hex is smaller than that of ring-1-7 only for very large SNRs (> 15 dB); in this range, the capacities of both schemes have already saturated at 3 bits/symbol.

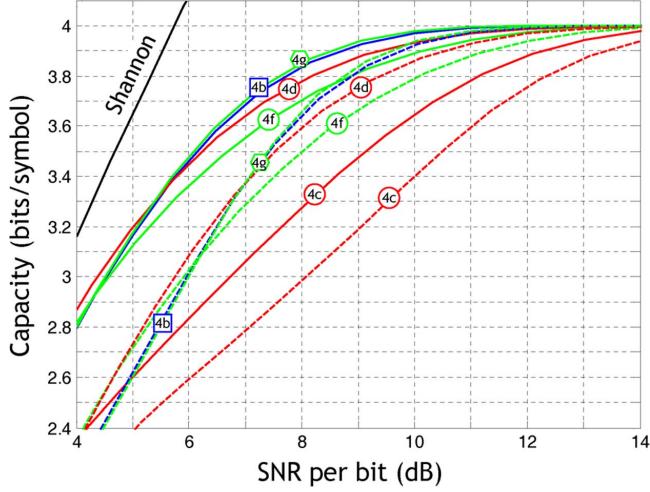


Fig. 46. Capacities of various 16-ary formats with soft (solid lines) and hard decision (dashed lines): hexagonal (4g), QAM (4b), PSK (4c), ring-1-3-5-7 (4f), 2-ASK/8-PSK (4d).

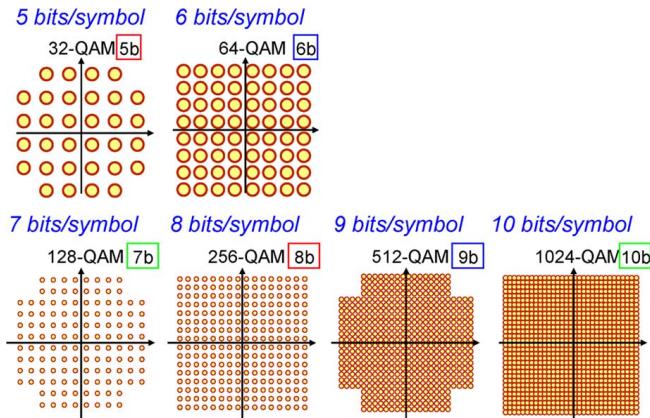


Fig. 47. QAM constellations with 32, 64, 128, 256, 512, and 1024 points.

C. 16-ary Constellations

The most prominent 16-ary constellations are (sorted by minimum distance): 16-hex (4g) ($d_{\min} = 0.669$), 16-QAM (4b) ($d_{\min} = 0.632$), 4-ASK/4-PSK (4e) ($d_{\min} = 0.516$), 2-ASK/8-PSK (4d) ($2d_{\min} = 0.484$), ring-1-3-5-7 (4f) ($d_{\min} = 0.431$), 16-PSK (4c) ($d_{\min} = 0.39$). Fig. 44 contains illustrations of these constellations.

The performance of selected 16-ary modulation schemes is depicted in Figs. 45 and 46, respectively. As expected, the modulation schemes perform according to their minimum distance at large SNRs. The hexagonal constellation achieves the highest capacity; however, the gain over 16-QAM is marginal. Because of the simpler implementation in practical systems, 16-QAM is generally preferred over 16-hex. Across the SNR range shown in Fig. 46, 2-ASK/8-PSK and 4-ASK/4-PSK achieve virtually the same capacity values. 16-PSK performs worst in terms of error ratio or capacity at all values of SNR shown. With increasing M , the gain obtained from soft decoding increases.

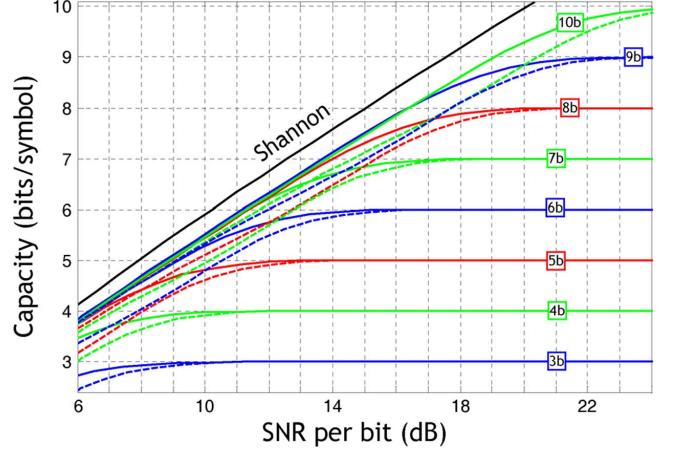


Fig. 48. Capacities of M -QAM constellations ($M = 8, 16, 32, 64, 128, 256, 512, 1024$) with soft (solid lines) and hard decision (dashed lines).

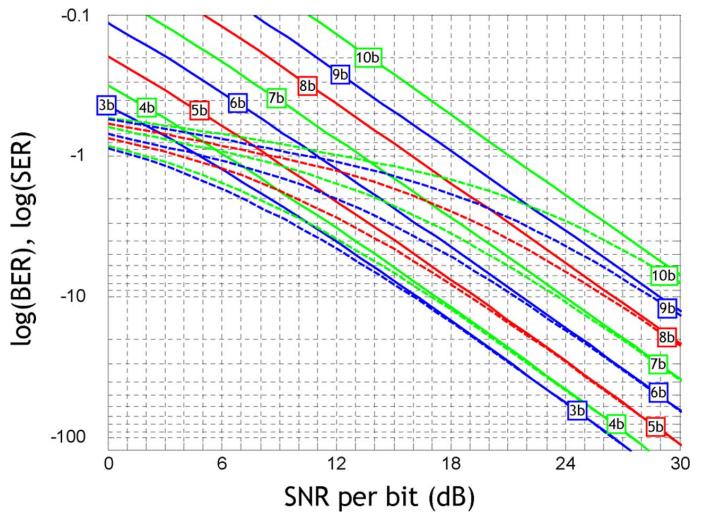


Fig. 49. SER (solid) and BER (dashed) curves for M -QAM constellations.

D. Higher Order Constellations

As M increases, the trend observed for $M \leq 16$ continues: At large SNRs, those constellations with the largest minimum distance achieve the highest capacity. These are hexagonal constellations, but QAM constellations perform only slightly worse. Ring constellations and ASK/PSK combinations are optimum at low SNRs. In absolute values, however, their capacity gain over QAM is very small. The SER/BER curves as well as soft- and hard-decision capacities for all M -QAM constellations with $M = 2^i, i = 3, \dots, 10$ are shown in Figs. 48 and 49, respectively.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge discussions with T. Freckmann, J. P. Gordon, B. Basch, R. W. Tkach, A. R. Chraplyvy, A. A. M. Saleh, M. Joindot, S. Korotky, M. Magarini, S. Colas, A. Gnauck, D. Caplan, P. Andrekson, G. P. Agrawal, A. Chowdhury, and H. Kogelnik, and many others not mentioned here.

TABLE VII
LIST OF ACRONYMS

Acronym	Explanation
ADC	Analog-to-digital converter
ASE	Amplified spontaneous emission
ASK	Amplitude-shift keying
AWGN	Additive white Gaussian noise
BER	Bit-error ratio
BPSK	Binary phase-shift keying
BR	Back-rotated
BWP	Backward pumps
ccG	Circularly symmetric complex Gaussian
CD	Chromatic dispersion
COI	Channel of interest
CPM	Continuous phase modulation
DAC	Digital-to-analog converter
dB	Decibel
DRB	Double Rayleigh backscattering or backscatter
DSP	Digital signal processing
ECOC	European conference on optical communication
EDFA	Erbium-doped fiber amplifier
ETDM	Electronically time-division multiplexed
FEC	Forward-error correction
FFT	Fast Fourier transform
FIR	Finite-impulse response
FWM	Four-wave mixing
FWP	Forward pumps
GVD	Group velocity dispersion
GNSE	Generalized non-linear Schrödinger equation
IDRA	Ideal distributed Raman amplification
I	In-phase
IFWM	Intra-channel four-wave mixing
ISI	Intersymbol-interference
IXPM	Intra-channel cross-phase modulation
LDPC	Low-density parity-check
LO	Local oscillator
MAP	Maximum a posteriori
MI	Modulation instability
ML	Maximum likelihood
MPI	Multi-path interference
M-PSK	M -ary phase-shift keying
MSK	Minimum-shift keying
N	Noise
NL	Nonlinearity
NRD	Net residual dispersion
OEO	Optical-electronic-optical
OF	Optical filter
OFC	Optical fiber communications conference
OFDM	Orthogonal frequency-division multiplexing
OOK	On/off keying
ORN	Optically-routed network
OSNR	Optical signal-to-noise ratio
PDF	Probability density function
PSK	Phase-shift keying
Q	Quadrature
QAM	Quadrature-amplitude modulation
QPSK	Quaternary phase or quadrature phase or quadri-phase phase-shift keying
r-ASK/PSK	r levels of amplitude-shift keying combined with continuous phase-shift keying (ring constellations)
RDPS	Residual dispersion per span
ROADM	Reconfigurable optical add/drop multiplexer
Rx	Receiver
SE	Spectral efficiency
SER	Symbol error ratio
SNR	Signal-to-noise ratio
SPDM	Singly-periodic dispersion map
SPM	Self-phase modulation
SBS	Stimulated Brillouin scattering
SRS	Stimulated Raman scattering
SSB	Single sideband
SSFM	Split-step Fourier method
SSMF	Standard single-mode fiber
Tx	Transmitter
WDM	Wavelength-division multiplexing
XPM	Cross-phase modulation

TABLE VIII
LIST OF SYMBOLS AND NOTATION

Notation	Explanation	Main occurrence
α	Fiber loss coefficient	Eq. (55)
α_{dB}	Fiber loss coefficient in dB	Eq. (55)
α_R	Power loss coefficient due to Rayleigh backscatter	Eq. (57)
$\Re\{a\}$	Real part of a	
a^*	Complex conjugate of a	
A_{eff}	Fiber effective area	Eq. (64)
β_2	Fiber group-velocity dispersion	Eq. (61)
B	WDM channel bandwidth	Fig. 3
B_e	Electrical bandwidth	Eq. (49)
B_{ref}	Reference bandwidth of 12.5 GHz	Eq. (33)
c	Speed of light in vacuum	
C	Capacity in bits per second	Eq. (24)
\tilde{C}	Capacity in bits per symbol	Eq. (13)
C_D^{\max}	Maximum cumulative dispersion	End of Sec. IX-C1
$\mathcal{E}[X]$	Expectation of X	Eq. (9)
$\Delta\phi$	Angle of rotation	Sec. X-C
$\delta(x)$	Dirac functional of x	Eq. (53)
d_{\min}	Smallest Euclidean distance between any two symbols of the constellation	App. XVI
D	Fiber chromatic dispersion	Eq. (61)
ε	Coupling coefficient	Eqs. (84)-(87) in App. XV
e	Elementary charge	
E	Pulse energy	
$\tilde{E}(z, \omega)$	Fourier transform of field $E(z, t)$	Eq. (62)
E_b	Energy per information bit	Below Eq. (31)
E_{in}	Input field	Sec. X-C
E_{LO}	Local oscillator field	Eq. (47)
E_{out}	Output field	Sec. X-C
E_s	Signal field	
f	Frequency	
f_c	Carrier frequency, $f_c > W/2$	
g	Distributed gain coefficient	Eq. (57)
h	Planck constant	
$H(f)$	Filter complex transfer function	Fig. 25
$h(t)$	Filter impulse response	App. XV
$H(X)$	Entropy of the random variable X	Eq. (8)
$H(X^n)$	Entropy of $X^n = X_1, X_2, \dots, X_n$	Eq. (42)
$H(X Y = b)$	Entropy of X conditioned on $Y = b$	Eq. (10)
$H(X Y)$	Entropy of X conditioned on Y	Eq. (11)
i	$\sqrt{-1}$	
$I(X; Y)$	Mutual information between X and Y	Eq. (12)
$I_0(\cdot)$	Modified Bessel function	Eq. (75) of App. XIV
k_B	Boltzmann constant	
K	Number of sub-channels	Sec. XII
K_T	Phonon occupancy factor	Eq. (56)
L	System or fiber length	
L	Number of symbols in a sub-channel	Sec. XII
$L_B(P)$	Spectral broadening length (power dependent)	Sec. X-B
L_D	Dispersion length	Sec. X-B
L_{NL}	Nonlinear length	Sec. X-B
L_{eff}	Effective fiber length	Eq. (67)
M	Modulation alphabet size, gives $\log_2(M)$ bits/symbol	Fig. 5
ν	Optical frequency	
ν_s	Optical frequency of the signal	
ν_p	Optical frequency of the Raman pump	
n	Number of optical filters	Fig. 25
n_2	Fiber nonlinear refractive index	Eq. (64)
n_{ij}	Number of bit errors made for a symbol error	Eq. (100) in App. XVI
n_{sp}	Spontaneous emission factor	Eq. (54)
N	Complex Gaussian noise random variable with variance σ_N^2	Eq. (14)
$N(z, t)$	Noise field dependent on distance z and time t	Eq. (53)
N_{ASE}	Amplified spontaneous emission spectral density, taken to be $N_{\text{ASE}} = N_0$	Eq. (33) and (53)
N_0	One-sided noise power spectral density (consider positive frequency only)	Eq. (24)
N_A	Number of optical amplifiers	Eq. (54)
N_I	Imaginary part of N	Below Eq.(14)
N_R	Real part of N	Below Eq.(14)

TABLE IX
LIST OF SYMBOLS AND NOTATION

OSNR	Optical signal-to-noise ratio	
ϕ, θ	Phase	
ϕ_{NL}	Integrated nonlinear phase	Eq. (68)
Φ_{NL}	Integrated nonlinear phase spectral density	Below Eq. (68)
Φ_{XPM}	Average phase rotation caused by XPM	Below Eq. (29)
ϕ	Angle	
p	Number of polarizations used for signaling	Eq. (34)
P	Power	
P_{DRB}	Power associated to double Rayleigh backscatter	Eq. (57)
P_m	Input signal power	
$p_X(a)$	Probability density function of X evaluated at $X = a$	Eq. (8)
$p_{XY}(a, b)$	Joint probability density function of $[X Y]$ evaluated at $[X Y] = [a b]$	Sec. III-A
$p_{Y X}(b a)$	Conditional probability density function of Y given X conditioned at $[X Y] = [a b]$	Sec. III-A
$P_X(a)$	Probability that $X = a$	Sec. III-A
$P_{XY}(a, b)$	Joint probability that $[X Y] = [a b]$	Sec. III-A
$P_{Y X}(b a)$	Conditional probability that $Y = b$ conditioned on $X = a$	Sec. III-A
Q	Receiver sampling alphabet size	Fig. 5
ρ	Raised-cosine spectrum roll-off factor	Eq. (6)
r	Amplitude	Sec. X-C
r	Number of rings	Fig. 15 and end of App. XIV
R_s	Transmit symbol rate, baud rate	Fig. 3
\mathcal{R}_x	Decision region corresponding to input symbol x	Eq. (99) in App. XVI
\bar{R}	Coded modulation rate in bits per symbol	Sec. III
\bar{R}_c	Code rate, unitless	Sec. III
SE	Spectral efficiency C/B	Eq. (26)
SNR	Signal-to-noise ratio	Below Eq. (23) and Sec. III-C
SNR_b	Signal-to-noise ratio per information bit	Eq. (31)
σ_{elec}^2	Variance of electronic noise	Sec. IX-A4
$s(t)$	Raised-cosine pulse	Eq. (7)
S	Dimensionless backscatter recapture fraction	Eq. (57)
S	Sub-channel	Sec. XII Sec. XII
$\{S_k\}_{k=1}^K$	Set of symbols of sub-channel S	Sec. XII
$S(f)$	Raised-cosine spectrum	Eq. (6)
S_D	Photodetector sensitivity	Eq. (47)
$S_{D,i}$	Photodetector sensitivity of detector i	App. XV
S_i	Electrical current at input i of a coherent detector	Fig. 39 in App. XV
S_I	Electrical current in-phase with local oscillator	Eq. (47)
S_Q	Electrical current in-quadrature with local oscillator	Eq. (47)
$\text{sinc}(t)$	Function with $\text{sinc}(t) = \sin(\pi t)/(\pi t)$	Fig. 2
σ_N^2	Variance of N , sometimes taken to be a power	
t	Time	
T	Symbol period $T = 1/R_s$	
ω_s	Optical angular frequency of signal	
W	Signal spectral support	Fig. 3
\mathcal{X}	Modulation alphabet, size is M if discrete	Sec. III
$x(t)$	Time-domain signal	Sec. II-A
$x_b(t)$	Complex baseband signal, $x_b(t) = x_c(t) + i x_s(t)$	Eq. (3)
$x_c(t)$	Projection of $x(t)$ onto $\sqrt{2} \cos(2\pi f_c t)$, in-phase component of $x(t)$	Eq. (3)
$x_s(t)$	Projection of $x(t)$ onto $\sqrt{2} \sin(2\pi f_c t)$, quadrature component of $x(t)$	Eq. (3)
$\{x_{b,k}\}$	Sequence of samples of $x_b(t)$, complex	Sec. II-A
$\{x_{c,k}\}$	Sequence of samples of $x_c(t)$, real	Sec. II-A
$\{x_{s,k}\}$	Sequence of samples of $x_s(t)$, real	Sec. II-A
$\{x_k\}$	Sequence of samples, real or complex	Sec. II-A
x^k	Finite-length sequence of samples x_1, x_2, \dots, x_k	Sec. III-A
$X(f)$	Fourier transform of $x(t)$	
X	Channel input random variable	Fig. 6
X_R	Real part of X	
X_I	Imaginary part of X	
\mathcal{Y}	Receiver sampling alphabet, size is Q if discrete	Sec. III
$y(t)$	Time-domain signal at receiver	Sec. II-C
$\{y_k\}$	Sequence of receiver samples, real or complex	Sec. II-C
Y	Channel output random variable	Fig. 6
Y_R	Real part of Y	
Y_I	Imaginary part of Y	
z	Distance	

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, p. 379–423 and 623–656, 1948.
- [2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968, ch. 2 and 4.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
- [4] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [5] J. M. Wozencraft and B. Reiffen, *Sequential Decoding*. New York: MIT Press and Wiley, 1961.
- [6] J. Massey, "Channel models for random-access systems," in *NATO Advances Studies Institutes Series E142*, 1988, pp. 391–402.
- [7] I. Karet and S. Shamai, "On the capacity of a twisted-wire pair: Gaussian model," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 379–383, Mar. 1990.
- [8] J. J. Werner, "The HDSL environment," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 6, pp. 785–800, Aug. 1991.
- [9] M. Gagnaire, "An overview of broadband access technologies," *Proc. IEEE*, vol. 85, no. 12, pp. 1958–1972, Dec. 1997.
- [10] A. Sendonaris, V. V. Veeravalli, and B. Aazhang, "Joint signaling strategies for approaching the capacity of twisted-pair channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 673–685, May 1998.
- [11] *The Communications Handbook*, J. D. Gibson, Ed. Boca Raton, FL: CRC Press, 1997.
- [12] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, pp. 41–59, 1996.
- [13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [14] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [15] A. F. Molisch, *Wireless Communications*. New York: Wiley, 2005.
- [16] J. P. Gordon, "Quantum effects in communications systems," *Proc. IRE*, vol. 50, pp. 1898–1908, 1962.
- [17] J. Pierce, "Optical channels: Practical limits with photon counting," *IEEE Trans. Commun.*, vol. COM-26, no. 12, pp. 1819–1821, Dec. 1978.
- [18] D. O. Caplan, "Laser communication transmitter and receiver design," *J. Opt. Fiber Commun. Res.*, vol. 4, no. 4, pp. 225–362, 2007.
- [19] E. T. Whittaker, "On the functions which are represented by the expansions of the interpolation theory," *Proc. R. Soc. Edinburgh*, vol. 35, pp. 181–194, 1915.
- [20] H. Nyquist, "Certain topics of telegraph transmission theory," *Trans. Amer. Inst. Electr. Eng.*, vol. 47, pp. 617–644, 1928.
- [21] A. Jerrí, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.
- [22] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
- [23] S. Haykin, *Communication Systems*, 4th ed. New York: Wiley, 2001.
- [24] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
- [25] B. Sklar, *Digital Communications*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.
- [26] R. G. Gallager, *Principles of Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [27] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [28] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed. New York: Macmillan, 2007.
- [29] R. G. Lyons, *Understanding Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 2004.
- [30] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. Commun.*, vol. COM-23, no. 9, pp. 921–934, Sep. 1975.
- [31] T. Aulin, N. Rydbeck, and C. E. Sundberg, "Continuous phase modulation—Part 2: Partial response signaling," *IEEE Trans. Commun.*, vol. COM-29, no. 3, pp. 210–225, Mar. 1981.
- [32] T. Aulin and C. E. Sundberg, "Continuous phase modulation—Part 1: Full response signaling," *IEEE Trans. Commun.*, vol. COM-29, no. 3, pp. 196–209, Mar. 1981.
- [33] J. B. Anderson, T. Aulin, and C. E. Sundberg, *Digital Phase Modulation*. New York: Plenum, 1986.
- [34] A. Lender, "Correlative level coding for binary data transmission," *IEEE Spectrum*, vol. 3, no. 2, pp. 104–115, Feb. 1966.
- [35] J. L. Massey, "The how and why of channel coding," in *Proc. Int. Zurich Semin.*, 1984, pp. 67–73.
- [36] S. Pasupathy, "Minimum shift keying: A spectrally efficient modulation," *IEEE Commun. Mag.*, vol. 17, no. 4, pp. 14–22, Jul. 1979.
- [37] J. Massey, "A generalized formulation of minimum shift keying modulation," in *Proc. IEEE Int. Conf. Commun.*, 1980, pp. 26.5.1–26.5.4.
- [38] B. E. Rimoldi, "A decomposition approach to CPM," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 260–270, Mar. 1988.
- [39] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2007.
- [40] S. Lin and D. J. Costello, *Error Control Coding*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2004.
- [41] R. H. Walden, "Performance trends for analog to digital converters," *IEEE Commun. Mag.*, vol. 37, no. 2, pp. 96–101, Feb. 1999.
- [42] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [43] R. H. Walden, "Analog-to-digital converters and associated IC technologies," in *Proc. IEEE Compound Semicond. Integr. Circuits Symp.*, 2008, pp. 1–2.
- [44] R. M. Gray, *Source Coding Theory*. New York: Springer-Verlag, 1989.
- [45] M. Bayes and M. Price, "An essay towards solving a problem in the doctrine of chances," *Philos. Trans. R. Soc. Lond.*, vol. 53, pp. 370–418, 1763.
- [46] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2002.
- [47] G. Kramer, A. Ashikhmin, A. J. van Wijngaarden, and X. Wei, "Spectral efficiency of coded phase-shift keying for fiber-optic communication," *J. Lightw. Technol.*, vol. 21, no. 10, pp. 2438–2445, Oct. 2003.
- [48] J. Hancock and R. Lucky, "Performance of combined amplitude and phase-modulated communication systems," *Commun. Syst., IRE Trans.*, vol. 8, no. 4, pp. 232–237, 1960.
- [49] J. M. Geist, "Capacity and cutoff rate for dense M-ary PSK constellations," in *Proc. IEEE Mil. Commun. Conf. Record*, 1990, pp. 768–770.
- [50] P. J. Winzer and R. J. Essiambre, "Advanced optical modulation formats," *Proc. IEEE*, vol. 94, no. 5, pp. 952–985, May 2006.
- [51] P. J. Winzer and R.-J. Essiambre, "Advanced optical modulation formats," in *Optical Fiber Telecommunications VB*, I. Kaminow and T. Li, Eds. New York: Academic, 2008, pp. 232–304.
- [52] F. Gray, "Pulse Code Communication," U.S. Patent 2 632 058, 1953.
- [53] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.
- [54] R. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Labs Tech. J.*, vol. 14, no. 4, pp. 3–9, 2010.
- [55] J. B. Stark, "Fundamental limits of information capacity for optical communications channels," in *Proc. Eur. Conf. Opt. Commun.*, 1999, pp. I–28.
- [56] E. Desurvire, "A quantum model for optically amplified nonlinear transmission systems," *Opt. Fiber Technol.*, vol. 8, pp. 210–230, 2002.
- [57] E. Desurvire, "A common quantum noise model for optical amplification and nonlinearity in WDM transmission," in *Proc. Eur. Conf. Opt. Commun.*, 2002, p. Tu3.1.1.
- [58] E. Desurvire, *Classical and Quantum Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [59] J. Tang, "The Shannon channel capacity of dispersion-free nonlinear optical fiber transmission," *J. Lightw. Technol.*, vol. 19, no. 8, pp. 1104–1109, Aug. 2001.
- [60] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, pp. 1027–1030, 2001.
- [61] E. Narimanov and P. P. Mitra, "The channel capacity of a fiber optics communication system: Perturbation theory," *J. Lightw. Technol.*, vol. 20, no. 3, pp. 530–537, Mar. 2002.
- [62] J. Tang, "The channel capacity of a multispot DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver," *J. Lightw. Technol.*, vol. 20, no. 7, pp. 1095–1101, Jul. 2002.
- [63] 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, 2003, Paper 203901.
- [64] L. G. L. Wegener, M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood, "The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: Cross-phase modulation and four-wave mixing," *Physica D*, vol. 189, pp. 81–99, 2004.
- [65] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in a nonlinear optical fiber: Coherent communication," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5008–5022, Nov. 2006.

- [66] K.-P. Ho, "Channel capacity of WDM systems using constant-intensity modulation formats," in *Proc. Opt. Fiber Commun. Conf.*, 2002, pp. 731–733.
- [67] I. 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.
- [68] M. Ivkovic, I. Djordjevic, and B. Vasic, "Calculation of achievable information rates of long-haul optical transmission systems using instanton approach," *J. Lightw. Technol.*, vol. 25, no. 5, pp. 1163–1168, May 2007.
- [69] I. Djordjevic, L. Xu, and T. Wang, "On the channel capacity of multilevel modulation schemes with coherent detection," in *Proc. Asia Commun. Photon. Conf. Exhib.*, 2009, p. ThC4.
- [70] I. Djordjevic, "Ultimate information capacity of fiber optic networks," *SPIE Photon. West*, vol. , no. 4, pp. 7619–7621, 2010.
- [71] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, G. Kramer, and E. C. Burrows, "The capacity of fiber-optic communication systems," in *Proc. Opt. Fiber Commun. Conf.*, 2008, p. OTuE1.
- [72] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," *Phys. Rev. Lett.*, vol. 101, 2008, Paper 163901.
- [73] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, and G. Kramer, "Exploring capacity limits of fibre-optic communication systems," in *Proc. Eur. Conf. Opt. Commun.*, 2008, Paper We1.E.1.
- [74] R.-J. Essiambre, G. Kramer, G. J. Foschini, and P. J. Winzer, "High spectral efficiency modulation for high capacity transmission," *Proc. IEEE/LEOS Summer Top. Meet.*, pp. 113–114, 2008.
- [75] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, and G. Kramer, "Capacity limits of fiber-optic communication systems," in *Proc. Opt. Fiber Commun. Conf.*, 2009, Paper OThL1.
- [76] H. Haunstein and M. Mayrock, "OFDM spectral efficiency limits from fiber and system non-linearities," in *Proc. Opt. Fiber Commun. Conf.*, 2010, Paper OTHM7.
- [77] O. A. Sab and V. Lemaire, "Block turbo code performances for longhaul DWDM optical transmission systems," in *Proc. Opt. Fiber Commun. Conf.*, 2000, pp. 280–282.
- [78] T. Mizuochi, "Recent progress in forward error correction and its interplay with transmission impairments," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, no. 4, pp. 544–554, Jul./Aug. 2006.
- [79] T. Mizuochi, "Next generation FEC for optical communication," in *Proc. Opt. Fiber Commun. Conf.*, 2008, Paper OTuE5.
- [80] D. F. Grosz, A. Agarwal, S. Banerjee, D. N. Maywar, and A. P. Küng, "All-Raman ultralong-haul single-wideband DWDM transmission systems with OADM capability," *J. Lightw. Technol.*, vol. 22, no. 2, pp. 423–432, Feb. 2004.
- [81] S. Radic, N. Vukovic, S. Chandrasekhar, A. Velingker, and A. Srivastava, "Forward error correction performance in the presence of rayleigh-dominated transmission noise," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 326–328, 2003.
- [82] P. J. Winzer, M. Pfennigbauer, and R.-J. Essiambre, "Coherent crosstalk in ultradense WDM systems," *J. Lightw. Technol.*, vol. 23, no. 4, pp. 1734–1744, Apr. 2005.
- [83] R.-J. Essiambre and P. J. Winzer, "Fibre nonlinearities in electronically pre-distorted transmission," in *Proc. European Conf. Opt. Commun.*, 2005, Paper Tu3.2.2.
- [84] R.-J. Essiambre, P. J. Winzer, X. Q. Wang, W. Lee, C. A. White, and E. C. Burrows, "Electronic predistortion and fiber nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 18, pp. 1804–1806, 2006.
- [85] P. J. Winzer, "Modulation and multiplexing in optical communication systems," LEOS Newsletter 2009 [Online]. Available: <http://www.ieee.org/organizations/pubs/newsletters/leos/feb09/index.html>
- [86] P. J. Winzer, G. Raybon, C. R. Doerr, M. Duelk, and C. Dorrer, "107-Gb/s optical signal generation using electronic time-division multiplexing," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 3107–3113, Jan. 2006.
- [87] T. Itoi, K. Fukuchi, and T. Kasamatsu, "Enabling technologies for 10 Tb/s transmission capacity and beyond," in *Proc. Eur. Conf. Opt. Commun.*, 2001, pp. 598–601.
- [88] F. Boubal, E. Brandon, L. Buet, S. Chernikov, V. Havard, A. Heerd, C. Hugbart, W. Idler, L. Labrunie, P. Le Roux, S. A. E. Lewis, A. Pham, L. Piriou, R. Uhel, and J.-P. Blondel, "Wideband and ultra-dense WDM transmission technologies toward over 10-Tb/s capacity," in *Proc. Eur. Conf. Opt. Commun.*, 2001, pp. 58–59.
- [89] K. Fukuchi, "Wideband and ultra-dense WDM transmission technologies toward over 10-Tb/s capacity," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2002, pp. 558–559.
- [90] *Optical Fiber Telecommunications V A + B*, I. Kaminow, T. Li, and A. E. Willner, Eds. New York: Academic, 2008.
- [91] A. R. Chraplyvy, "Limitations on lightwave communications imposed by optical-fiber nonlinearities," *J. Lightw. Technol.*, vol. 8, no. 10, pp. 1548–1557, Oct. 1990.
- [92] I. P. Kaminow, C. R. Doerr, C. Dragone, T. Koch, U. Koren, A. A. M. Saleh, A. J. Kirby, C. M. Özveren, B. Schofield, R. E. Thomas, R. A. Bany, D. M. Castagnozzi, V. W. S. Chan, B. R. Hemenway, D. Marquis, S. A. Parikh, M. L. Stevens, E. A. Swanson, S. G. Finn, and R. G. Gallager, "A wideband all-optical WDM network," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 5, pp. 780–799, Jun. 1996.
- [93] E. E. B. Basch, R. Egorov, S. Gringeri, and S. Elby, "Architectural tradeoffs for reconfigurable dense wavelength-division multiplexing systems," *IEEE J. Select. Topics Quantum Electron.*, vol. 12, no. 4, pp. 615–626, Jul./Aug. 2006.
- [94] R. Ramaswami and K. Sivarajan, *Optical Networks: A Practical Perspective*. San Mateo, CA: Morgan Kaufmann, 2001.
- [95] B. Mukherjee, *Optical WDM Networks*. New York: Springer-Verlag, 2006.
- [96] J. Simmons, *Optical Network Design and Planning*. New York: Springer-Verlag, 2008.
- [97] R.-J. Essiambre, B. Mikkelsen, and G. Raybon, "Intra-channel cross-phase modulation and four-wave mixing in high-speed TDM systems," *IEE Electron. Lett.*, vol. 35, pp. 1576–1578, 1999.
- [98] R.-J. Essiambre, G. Raybon, and B. Mikkelsen, , I. Kaminow and T. Li, Eds., "Pseudo-linear transmission of high-speed TDM signals: 40 and 160 Gb/s," in *Optical Fiber Telecommunications IV*. New York: Academic, 2002, pp. 232–304.
- [99] M. van Deventer, *Fundamentals of Bidirectional Transmission Over a Single Optical Fibre*. Norwell, MA: Kluwer, 1996.
- [100] C. Kim and Y. Chung, "2.5 Gb/s × 16-channel bidirectional WDM transmission system using bidirectional erbium-doped fiber amplifier based on spectrally interleaved synchronized etalon filters," *IEEE Photon. Technol. Lett.*, vol. 11, no. 6, pp. 745–747, 1999.
- [101] S. Radic and S. Chandrasekhar, "Limitations in dense bidirectional transmission in absence of optical amplification," *IEEE Photon. Technol. Lett.*, vol. 14, no. 1, pp. 95–97, 2002.
- [102] S. Radic, S. Chandrasekhar, P. Bernasconi, J. Centanni, C. Abraham, N. Copner, and K. Tan, "Feasibility of hybrid Raman/EDFA amplification in bidirectional optical transmission," *IEEE Photon. Technol. Lett.*, vol. 14, no. 2, pp. 221–223, 2002.
- [103] J. Ko, S. Kim, J. Lee, S. Won, Y. Kim, and J. Jeong, "Estimation of performance degradation of bidirectional WDM transmission systems due to Rayleigh backscattering and ASE noises using numerical and analytical models," *J. Lightw. Technol.*, vol. 21, no. 4, pp. 938–946, Apr. 2003.
- [104] P. J. Winzer, "Optical transmitters, receivers, and noise," in *Wiley Encyclopedia of Telecommunications*, J. G. Proakis, Ed. New York: Wiley, 2002, pp. 1824–1840 [Online]. Available: http://www.mrw.interscience.wiley.com/eot/eot_sample_fs.html, online
- [105] G. Einarsson, *Principles of Lightwave Communications*. New York: Wiley, 1996.
- [106] P. Henry, "Error-rate performance of optical amplifiers," in *Proc. Opt. Fiber Commun. Conf.*, 1989, vol. 5, Paper THK3.
- [107] D. Marcuse, "Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 8, no. 12, pp. 1816–1823, Dec. 1990.
- [108] J. Lee and C. Shim, "Bit-error-rate analysis of optically preamplified receivers using an eigenfunction expansion method in optical frequency domain," *J. Lightw. Technol.*, vol. 12, no. 7, pp. 1224–1229, Jul. 1994.
- [109] P. J. Winzer, "Performance estimation of receivers corrupted by optical noise," *Proc. Opt. Amplifiers Appl.*, vol. 60, pp. 268–273, 2001.
- [110] T. Li and M. Teich, "Bit-error rate for a lightwave communication system incorporating an erbium-doped fibre amplifier," *IEE Electron. Lett.*, vol. 27, pp. 598–600, 1991.
- [111] P. J. Winzer, S. Chandrasekhar, and H. Kim, "Impact of filtering on RZ-DPSK reception," *IEEE Photon. Technol. Lett.*, vol. 15, no. 6, pp. 840–842, 2003.
- [112] A. H. Gnauck and P. J. Winzer, "Optical phase-shift-keyed transmission," *J. Lightw. Technol.*, vol. 23, no. 1, pp. 115–130, Jan. 2005.
- [113] X. Liu, S. Chandrasekhar, and A. Leven, "Self-Coherent Optical Transport Systems," in *Optical Fiber Telecommunications VB*, I. Kaminow and T. Li, Eds. New York: Academic, 2008, pp. 131–178.
- [114] W. Schottky, "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern," *Annalen der Physik*, vol. 57, pp. 541–567, 1918.

- [115] J. Pierce, "Physical sources of noise," *Proc. IRE*, vol. 44, no. 5, pp. 601–608, 1956.
- [116] H. A. Haus, *Electromagnetic Noise and Quantum Optical Measurements*. New York: Springer-Verlag, 2000.
- [117] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, 2nd ed. New York: Wiley, 2007.
- [118] B. E. A. Saleh, *Photoelectron Statistics: With Applications to Spectroscopy and Optical Communication*. New York: Springer-Verlag, 1978.
- [119] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
- [120] P. J. Winzer, "Linking equations between photon statistics and photocurrent statistics for time-varying stochastic photon rates," *Quantum Semiclass. Opt.: J. Eur. Opt. Soc. Part B*, vol. 10, pp. 643–655, 1998.
- [121] E. Säckinger, *Broadband Circuits for Optical Fiber Communication*. New York: Wiley, 2005.
- [122] S. Alexander, *Optical Communication Receiver Design*. Herts, U.K.: Institution of Engineering and Technology, 1997.
- [123] A. Einstein, "On the quantum theory of radiation," *Phys. Zeits.*, vol. 18, p. 121, 1917.
- [124] C. W. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods With Applications to Quantum Optics*, 3rd ed. New York: Springer-Verlag, 2004.
- [125] E. Desurvire, D. Bayart, B. Destheux, and S. Bigo, *Erbium-Doped Fiber Amplifiers and Device and System Developments*. New York: Wiley, 2002.
- [126] P. Diament and M. C. Teich, "Evolution of the statistical properties of photons passed through a traveling-wave laser amplifier," *IEEE J. Quantum Electron.*, vol. 28, no. 5, pp. 1325–1334, May 1992.
- [127] T. Li and M. C. Teich, "Photon point process for traveling-wave laser amplifiers," *IEEE J. Quantum Electron.*, vol. 29, no. 9, pp. 2568–2578, Sep. 1993.
- [128] J. P. Gordon, W. H. Louisell, and L. R. Walker, "Quantum fluctuations and noise in parametric processes II," *Phys. Rev.*, vol. 129, pp. 481–485, 1963.
- [129] J. P. Gordon, L. R. Walker, and W. H. Louisell, "Quantum statistics of masers and attenuators," *Phys. Rev.*, vol. 130, pp. 806–812, 1963.
- [130] J. R. Barry, D. G. Messerschmitt, and E. A. Lee, *Digital Communication*, 5th ed. New York: Springer-Verlag, 2003.
- [131] A. Mecozzi, "Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers," *J. Lightw. Technol.*, vol. 12, no. 11, pp. 1993–2000, Nov. 1994.
- [132] C. W. Gardiner, *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Science*, 3rd ed. New York: Springer-Verlag, 2004.
- [133] P. C. Becker, N. A. Olsson, and J. R. Simpson, *Erbium-Doped Fiber Amplifiers, Fundamentals and Technology*. New York: Academic, 1999.
- [134] A. Bjarklev, *Optical Fiber Amplifiers: Design and System Applications*. Boston, MA and London: Artech House, 1993.
- [135] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [136] R. H. Stolen and E. P. Ippen, "Raman gain in glass optical waveguides," *Appl. Phys. Lett.*, vol. 22, pp. 276–278, 1973.
- [137] J. Bromage, "Raman amplification for fiber communications systems," *J. Lightw. Technol.*, vol. 22, no. 1, pp. 79–93, Jan. 2004.
- [138] V. E. Perlin and H. G. Winful, "On trade-off between noise and nonlinearity in WDM systems with distributed Raman amplification," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2002, pp. 178–180, paper WB1.
- [139] V. E. Perlin and H. G. Winful, "Optimizing the noise performance of broadband WDM systems with distributed Raman amplification," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1199–1201, 2002.
- [140] L. F. Mollenauer and J. P. Gordon, *Solitons in Optical Fibers: Fundamentals and Applications*. New York: Academic, 2006.
- [141] J.-C. Bouteiller, K. Brar, J. Bromage, S. Radic, and C. Headley, "Dual-order Raman pump," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 212–214, 2003.
- [142] T. J. Ellingham, J. D. Ania-Castanón, R. Ibbotson, X. Chen, L. Zhang, and S. K. Turitsyn, "Quasi-lossless optical links for broadband transmission and data processing," *IEEE Photon. Technol. Lett.*, vol. 18, pp. 268–270, 2006.
- [143] P. B. Hansen, L. Eskildsen, A. J. Stentz, T. A. Strasser, J. Judkins, J. J. DeMarco, R. Pedrazzani, and D. J. DiGiovanni, "Rayleigh scattering limitations in distributed Raman pre-amplifiers," , pp. 343–345, 1997.
- [144] S. A. E. Lewis, S. V. Chernikov, and J. R. Taylor, "Characterization of double Rayleigh scatter noise in Raman amplifiers," *IEEE Photon. Technol. Lett.*, vol. 12, no. 5, pp. 528–530, 2000.
- [145] R. Boyd, *Nonlinear Optics*. New York: Academic, 2008.
- [146] J. Bromage, P. J. Winzer, and R.-J. Essiambre, , M. N. Islam, Ed., "Multiple-path interference and its impact on system design," in *Raman Amplifiers and Oscillators in Telecommunications*. New York: Springer-Verlag, 2003.
- [147] M. Nissov, K. Rottwitt, H. Kidorf, and M. Ma, "Rayleigh crosstalk in long cascades of distributed unsaturated Raman amplifiers," *IEE Electron. Lett.*, vol. 35, pp. 997–998, 1999.
- [148] G. P. Agrawal, *Lightwave Technology: Components and Devices*. New York: Wiley, 2004.
- [149] D. Gloge, "Dispersion in weakly guiding fibers," *Appl. Opt.*, vol. 10, no. 11, pp. 2442–2445, 1971.
- [150] *Optical-Fiber Transmission*, E. E. B Basch, Ed. Indianapolis, IN: Sams Technical Publishing, 1986.
- [151] *Optical Fiber Telecommunications II*, S. Miller and I. Kaminow, Eds. New York: Academic, 1988.
- [152] Y. Akasaka, R. Sugizaki, A. Umeda, T. Kamiya, F. Ltd, and J. Chiba, "High-dispersion-compensation ability and low nonlinearity of w-shaped DCF," in *Proc. Opt. Fiber Commun. Conf.*, 1996, pp. 201–202.
- [153] A. Vengsarkar, "Dispersion compensating fibers," in *Proc. Opt. Fiber Commun. Conf.*, 1997, pp. 233–234.
- [154] L. Grüner-Nielsen, S. Knudsen, B. Edvold, T. Veng, D. Magnusson, C. Larsen, and H. Damsgaard, "Dispersion compensating fibers," *Opt. Fiber Technol.*, vol. 6, no. 2, pp. 164–180, 2000.
- [155] L. Grüner-Nielsen, M. Wandel, P. Kristensen, C. Jørgensen, L. Jørgensen, B. Edvold, B. Þálsdóttir, and D. Jakobsen, "Dispersion-compensating fibers," *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3566–3579, Jan. 2005.
- [156] M. Nishimura, "Optical fibers and fiber dispersion compensators for high-speed optical communication," *J. Opt. Fiber Commun. Rep.*, vol. 2, pp. 115–139, 2005.
- [157] P. Nouchi, L. de Montmorillon, P. Sillard, and A. Bertaina, "Advance in long-haul fibers," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2003, pp. 154–155, Paper TuB5.
- [158] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. San Diego, CA: Elsevier, 2006.
- [159] R. Lundin, "Dispersion flattening in a w fiber," *Appl. Opt.*, vol. 33, no. 6, pp. 1011–1014, 1994.
- [160] D. Marcuse, "Single-channel operation in very long nonlinear fibers with optical amplifiers at zero dispersion," *J. Lightw. Technol.*, vol. 9, no. 3, pp. 356–361, Mar. 1991.
- [161] D. Marcuse, A. R. Chraplyvy, and R. W. Tkach, "Effect of fiber nonlinearity on long-distance transmission," *J. Lightw. Technol.*, vol. 9, no. 1, pp. 121–128, Jan. 1991.
- [162] L. K. Wickham, R.-J. Essiambre, A. H. Gnauck, P. J. Winzer, and A. R. Chraplyvy, "Bit pattern length dependence of intrachannel nonlinearities in pseudolinear transmission," *IEEE Photon. Technol. Lett.*, vol. 16, pp. 1591–1593, 2004.
- [163] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-hall, 2002.
- [164] A. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
- [165] J. Kerr, "A new relation between electricity and light: Dielectrified media birefringent," *Phil. Mag.*, vol. 50, no. 19, p. 3337, 1875.
- [166] G. P. Agrawal, *Lightwave Technology: Telecommunication Systems*. New York: Wiley, 2005.
- [167] E. Ippen and R. Stolen, "Stimulated Brillouin scattering in optical fibers," *Appl. Phys. Lett.*, vol. 21, pp. 539–541, 1972.
- [168] R. Smith, "Optical power handling capacity of low loss optical fibers as determined by stimulated Raman and brillouin scattering," *Appl. Opt.*, vol. 11, no. 3, pp. 2489–2494, 1972.
- [169] R. H. Stolen, *Optical Fiber Telecommunications*. New York: Academic, 1979, ch. Nonlinear Properties of Optical Fibers, pp. 125–150.
- [170] A. R. Chraplyvy and P. S. Henry, "Performance degradation due to stimulated Raman scattering in wavelength-division-multiplexed optical-fibre systems," *IEE Electron. Lett.*, vol. 19, pp. 641–642, 1983.
- [171] R. Tkach, A. Chraplyvy, and R. Derosier, "Spontaneous Brillouin scattering for single-mode optical-fibre characterisation," *IEE Electron. Lett.*, vol. 22, pp. 1011–1013, 1986.
- [172] N. Bloembergen, "The stimulated Raman effect," *Amer. J. Phys.*, vol. 35, pp. 989–1023, 1967.

- [173] R. Hellwarth, J. Cherlow, and T. T. Yang, "Origin and frequency dependence of nonlinear optical susceptibilities of glasses," *Phys. Rev. B*, vol. 11, no. 2, pp. 964–967, 1975.
- [174] R. Stolen, "Nonlinearity in fiber transmission," *Proc. IEEE*, vol. 68, no. 10, pp. 1232–1236, Oct. 1980.
- [175] R. Stolen, J. Gordon, W. Tomlinson, and H. Haus, "Raman response function of silica-core fibers," *J. Opt. Soc. Amer. B*, vol. 6, no. 6, pp. 1159–1166, 1989.
- [176] K. Blow and D. Wood, "Theoretical description of transient stimulated Raman scattering in optical fibers," *IEEE J. Quantum Electron.*, vol. 25, no. 12, pp. 2665–2673, Dec. 1989.
- [177] P. Mamyshev and S. Chernikov, "Ultrashort-pulse propagation in optical fibers," *J. Opt. Soc. Amer. B*, vol. 15, no. 19, pp. 1076–1078, 1990.
- [178] S. Chernikov and P. Mamyshev, "Femtosecond soliton propagation in fibers with slowly decreasing dispersion," *J. Opt. Soc. Am. B*, vol. 8, no. 8, pp. 1633–1641, 1991.
- [179] L. Eskildsen, P. Hansen, P. Koren, B. Miller, M. Young, and K. Dreyer, "Stimulated Brillouin scattering suppression with low residual AM using a novel temperature wavelength-dithered DFB laser diode," *IEE Electron. Lett.*, vol. 32, no. 15, pp. 1387–1389, 1996.
- [180] M. Van Deventer and A. Boot, "Polarization properties of stimulated Brillouin scattering in single-mode fibers," *J. Lightw. Technol.*, vol. 12, no. 4, pp. 585–590, Apr. 1994.
- [181] D. Christodoulides and R. Jander, "Evolution of stimulated Raman crosstalk in wavelength division multiplexed systems," *IEEE Photon. Technol. Lett.*, vol. 8, no. 12, pp. 1722–1724, 1996.
- [182] M. Zirngibl, "Analytical model of Raman gain effects in massive wavelength division multiplexed transmission systems," *IEE Electron. Lett.*, vol. 34, pp. 789–790, 1998.
- [183] A. Chraplyvy, J. Nagel, and R. Tkach, "Equalization in amplified WDM lightwave transmission systems," *IEEE Photon. Technol. Lett.*, vol. 4, no. 8, pp. 920–922, 1992.
- [184] A. Tomita, "Cross talk caused by stimulated Raman scattering in single-mode wavelength-division multiplexed systems," *Opt. Lett.*, vol. 8, no. 7, pp. 412–414, 1983.
- [185] W. Jiang and P. Ye, "Crosstalk in fiber Raman amplification for WDM systems," *J. Lightw. Technol.*, vol. 7, no. 9, pp. 1407–1411, Sep. 1989.
- [186] F. Forghieri, R. W. Tkach, and A. R. Chraplyvy, "Effect of modulation statistics on Raman crosstalk in WDM systems," *IEEE Photon. Technol. Lett.*, vol. 7, no. 3, pp. 101–103, 1995.
- [187] *Raman Amplifiers for Telecommunications 1: Physical Principles*, M. Islam, Ed. New York: Springer-Verlag, 2004.
- [188] C. Headley and G. P. Agrawal, *Raman Amplification in Fiber Optical Communication Systems*. New York: Academic, 2005.
- [189] R.-J. Essiambre, P. J. Winzer, J. Bromage, and C. H. Kim, "Design of bidirectionally pumped fiber amplifiers generating double Rayleigh backscattering," *IEEE Photon. Technol. Lett.*, vol. 14, no. 7, pp. 914–916, 2002.
- [190] H. A. Haus, "The noise figure of optical amplifiers," *IEEE Photon. Technol. Lett.*, vol. 10, no. 11, pp. 1602–1604, 1998.
- [191] H. J. Landau, "On the recovery of a band-limited signal, after instantaneous companding and subsequent band limiting," *Bell Syst. Tech. J.*, vol. 39, pp. 351–364, 1960.
- [192] H. J. Landau and W. L. Miranker, "The recovery of distorted bandlimited signals," *J. Math Anal. Appl.*, vol. 2, pp. 97–104, 1961.
- [193] S. Bigo, "Modelling of WDM terrestrial and submarine links for the design of WDM networks," in *Proc. Opt. Fiber Commun. Conf.*, 2006, Paper OThD1.
- [194] A. Mecozi, C. B. Clausen, M. Shtaif, S.-G. Park, and A. H. Gnauck, "Cancellation of timing and amplitude jitter in symmetric links using highly dispersed pulses," *IEEE Photon. Technol. Lett.*, vol. 13, pp. 445–447, 2001.
- [195] T. Freckmann, R.-J. Essiambre, P. J. Winzer, G. J. Foschini, and G. Kramer, "Fiber capacity limits with optimized ring constellations," *IEEE Photon. Technol. Lett.*, vol. 21, pp. 1496–1498, 2009.
- [196] M. Eiselt, "Does spectrally periodic dispersion compensation reduce non-linear effects?" in *Proc. European Conf. Opt. Commun.*, 1999, vol. 1, pp. 144–145.
- [197] G. Bellotti and S. Bigo, "Cross-phase modulation suppressor for multispans dispersion-managed WDM transmissions," *IEEE Photon. Technol. Lett.*, vol. 12, no. 6, pp. 726–728, 2000.
- [198] G. Bellotti, S. Bigo, S. Gauchard, P. Cortes, and S. LaRochelle, "10 × 10 gb/s cross-phase modulation suppressor using WDM narrowband fiber Bragg gratings," *IEEE Photon. Technol. Lett.*, vol. 12, pp. 1403–1405, 2000.
- [199] C. Xie, "Suppression of inter-channel nonlinearities in WDM coherent PDM-QPSK systems using periodic-group-delay dispersion compensators," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Vienna, Austria, 2009, p. Paper P4.08.
- [200] J. Yu, X. Zhou, M.-F. Huang, Y. Shao, D. Qian, T. Wang, M. Cvijetic, P. Magill, L. Nelson, M. Birk, S. Ten, H. B. Matthew, and S. K. Mishra, "17 Tb/s (161 114 Gb/s) PolMux-RZ-8PSK transmission over 662 km of ultra-low loss fiber using C-band EDFA amplification and digital coherent detection," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2008, Paper Th3.E.2.
- [201] H. Takahashi, A. A. Amin, S. L. Jansen, I. Morita, and H. Tanaka, "8 × 66.8-Gbit/s coherent PDM-OFDM transmission over 640 km of SSMF at 5.6-bit/s/Hz spectral efficiency," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2008, Paper Th3.E.4.
- [202] P. J. Winzer and A. H. Gnauck, "112-Gb/s polarization-multiplexed 16-QAM on a 25-GHz WDM grid," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2008, Paper Th3.E.5.
- [203] H. Takahashi, A. A. Amin, S. L. Jansen, I. Morita, and H. Tanaka, "DWDM transmission with 7.0 bit/s/Hz spectral efficiency using 8 × 65.1 Gbit/s coherent PDM OFDM signals," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2009, Paper PDPB7.
- [204] A. H. Gnauck, P. J. Winzer, C. R. Doerr, and L. L. Buhl, "10 112-Gb/s PDM 16-QAM transmission over 630 km of fiber with 6.2-b/s/Hz spectral efficiency," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2009, Paper PDDB8.
- [205] P. J. Winzer and A. Kalmar, "Sensitivity enhancement of optical receivers by impulsive coding," *J. Lightw. Technol.*, vol. 17, no. 2, pp. 171–177, Feb. 1999.
- [206] N. A. Olsson, "Lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 7, no. 7, pp. 1071–1082, Jul. 1989.
- [207] J. Forney, Jr., G. D. , and L.-F. Wei, "Multidimensional constellations—Part 1: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.
- [208] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.
- [209] B. W. Kernighan and S. Lin, "Heuristic solution of a signal design optimization problem," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1145–1159, 1973.
- [210] G. F. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two dimensional signal constellations in the presence of Gaussian noise," *IEEE Trans. Commun.*, vol. 22, no. 1, pp. 28–38, Jan. 1974.
- [211] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. SAC-2, no. 5, pp. 632–647, Sep. 1984.
- [212] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-Carrier Quadrature Amplitude Modulation*. New York: Wiley, 2000.
- [213] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 2nd ed. New York: Springer-Verlag, 1993.



René-Jean Essiambre (S'95–M'06) received the B.S. and Ph.D. degrees in physics and optics from Université Laval, Quebec City, QC, Canada.

During his Ph.D. studies, he spent one year at McGill University, Montreal, QC, Canada, where he was engaged in research on solid-state physics. From 1995 to 1997, he was a Postdoctoral Fellow with Prof. Agrawal at The Institute of Optics, University of Rochester, Rochester, NY. Since 1997, he has been at Bell Laboratories, Alcatel-Lucent, Holmdel, NJ. His research has been focused on optical switching, soliton communication systems, high-power fiber lasers, and mode-locked fiber lasers. His current research interests include high-speed transmission (100 Gb/s and above) and physical layer design of fiber-optic communication systems, including Raman amplification, Rayleigh backscattering, fiber nonlinearities, network design, advanced modulation formats, information theory, and coding. He is the author and coauthor of more than 75 scientific publications and several book chapters. He has served on many conference committees including ECOC, OFC, CLEO, and LEOS.

Dr. Essiambre is a Fellow of the Optical Society of America (OSA). He is the recipient of the 2005 OSA Engineering Excellence Award. He is also a Distinguished Member of Technical Staff at Bell Laboratories.



Gerhard Kramer (S'91–M'94–SM'08–F'10) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Manitoba, MB, Canada in 1991 and 1992, respectively, and the Dr. sc.techn. (Doktor der technischen Wissenschaften) degree from the ETH Zurich, Zurich, Switzerland, in 1998.

From 1998 to 2000, he was with Endora Tech AG, Basel, Switzerland. From 2000 to 2008, he was with Bell Labs, Alcatel-Lucent, Murray Hill, NJ. Since 2009, he has been with the University of Southern California, Los Angeles. His research interests include information theory, coding and modulation, iterative decoding, and cooperative communications.

Dr. Kramer has been elected to the Board of Governors of the IEEE Information Theory Society for 2009–2011. He has served as Associate Editor, Special Issue Guest Editor, and Publications Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. He was technical program Cochair of the 2008 IEEE International Symposium on Information Theory, and general Cochair of the 1st and 2nd Annual Schools of Information Theory in 2008 and 2009. He is serving as a Member of the Emerging Technologies Committee of the IEEE Communications Society. He is a corecipient of the IEEE Communications Society 2005 Stephen O. Rice Prize Paper Award, a corecipient of the 2003 Bell Labs President's Gold Award, and a recipient of an ETH Medal in 1998.



Gerard J. Foschini (S'60–M'62–SM'83–F'86) received the B.S.E.E. degree from New Jersey Institute of Technology, Newark, the M.E.E. degree from New York University, Manhattan, and the Ph.D. degree in mathematics from Stevens Institute of Technology, Hoboken, NJ.

He has been at Bell Laboratories, Alcatel-Lucent, Holmdel, NJ, for over four decades, where he is currently a Distinguished Inventor in the Wireless Research Laboratory at Crawford Hill, Holmdel. He has done extensive research on point-to-point communication systems as well as on networks. He was also engaged in teaching at Princeton University and is currently a Visiting Professor in the graduate faculty in the Electrical Engineering Department, Rutgers University, Piscataway, NJ. He has published more than 100 papers and holds 15 patents.

Dr. Foschini is a Fellow of Bell Laboratories. He is the recipient of the 2000 Bell Labs Inventor's Award, for one of his wireless inventions on multiple antenna communications, and the 2002 Thomas Alva Edison Patent Award. He has also won the Bell Labs President's Gold Award "for the outstanding level of innovation, technical excellence and business impact." He also won the Bell Labs Team Award, the 2006 IEEE Eric E. Sumner Award, and the 2008 Alexander Graham Bell Medal for his work on multilevel antenna technology for high spectral-efficiency communications. He has been currently elected to the National Academy of Engineering.



Peter J. Winzer (S'93–A'99–SM'05–F'09) received the Ph.D. degree in electrical/communications engineering from the Vienna University of Technology, Vienna, Austria, in 1998.

His academic work, largely supported by the European Space Agency (ESA), was related to the analysis and modeling of space-borne Doppler wind lidar and highly sensitive free-space optical communication systems. In this context, he specialized on advanced digital optical modulation formats and high-sensitivity optical receivers using coherent and direct detection. After joining Bell Labs in November 2000, he focused on various aspects of high-bandwidth optical communication networks, including Raman amplification, optical modulation formats, advanced optical receiver concepts, and digital signal processing at bit rates from 10 to 100-Gb/s. He has widely published in peer-reviewed journals and at conferences and holds several patents in the fields of optical communications, lidar, and data networking.

Dr. Winzer is actively involved as a Reviewer, Associate Editor, and Committee Member of various journals and conferences and serves as an elected member of the IEEE-LEOS BoG. He is a Distinguished Member of technical Staff at Bell Labs, a Member of the Optical Society of America (OSA).



Bernhard Goebel (S'04) received the Diplom-Ingenieur degree in electrical engineering and information technology from Technische Universität München (TUM), Munich, Germany, in 2004. He is currently working toward the Dr.-Ing. degree at TUM. His diploma thesis was on the information-theoretic analysis of complex genetic diseases.

He was at the University of Southampton, Southampton, U.K., and Siemens Corporate Research, Princeton, NJ, during graduate semesters, where his research was focused on segmentation algorithms for medical imaging. Since 2004, he has been a Member of the technical staff at the Institute for Communications Engineering, TUM. During 2009, he was a visiting Researcher for three months at Alcatel-Lucent Bell Laboratories, Holmdel, NJ. His current research interests include information-theoretic aspects of optical communication systems. He has published in the fields of medical imaging, bioinformatics, optical communications and information theory, and is the inventor of one U.S. patent.

Mr. Goebel is a student member of the IET and VDE/ITG.