



CHALMERS

Chalmers Publication Library

Capacity bounds in optical communications [Invited tutorial]

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

European Conference on Optical Communication (ECOC)

Citation for the published paper:

Agrell, E. (2017) "Capacity bounds in optical communications [Invited tutorial]". European Conference on Optical Communication (ECOC)

Downloaded from: <http://publications.lib.chalmers.se/publication/254749>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Capacity Bounds in Optical Communications

Erik Agrell

Department of Electrical Engineering, Chalmers University of Technology, Sweden, agrell@chalmers.se

Abstract *The fundamental concept of channel capacity and related information-theoretic metrics are reviewed and some techniques to quantify them by means of lower and upper bounds are explained. The importance of properly discretizing the channel model is highlighted.*

Introduction: What is capacity?

Claude Shannon proved in 1948 that no matter how badly a communication channel distorts a transmitted signal, it is always possible to transmit digital information across it almost without errors, under the condition that the data rate is not too high¹. The transition between sufficiently low and too high data rate is quite abrupt², and the rate where this happens is in information theory defined as the *channel capacity*. In other words, the capacity is the *maximum achievable rate over a given channel for virtually error-free transmission*.

The *channel* includes not only the physical transmission medium between the source and the destination, but also often parts of the transmitter and receiver; namely, those parts that the designer is unwilling or unable to change³. For a mathematical treatment, a statistical description is needed of the output signal for a given input signal. We distinguish between *continuous-time* channels, for which the transmitted signals are waveforms, and *discrete-time* channels, for which the transmitted signals are sequences of samples. The relation between the input and output signals may in the former case be described by, e.g., a stochastic differential equation and in the latter case by a conditional probability distribution.

The maximization should be done over all parts of the transmitter and receiver that are not included in the channel model. This may include modulation, pulse shaping, filtering, sampling, demodulation, detection, forward error correction (FEC) coding, and other kinds of signal processing. Additional constraints may apply, for example if the capacity is to be quantified as a function of power and/or bandwidth.

The data rate, and therefore the capacity, can be normalized in different ways, and some confusion unfortunately exists between different versions of capacity. A safe way to circumvent the potential ambiguity is by giving the unit whenever capacity is discussed. The capacity C_{dt} of discrete-time channels is measured in bit/channel

use and the capacity C of continuous-time channels is measured in bit/s. Normalizing this quantity by the signal bandwidth W yields the *spectral efficiency* $SE = C/W$, which is measured in bit/s/Hz. Researchers with experience from linear systems should be warned that the capacity in bit/channel use is *not* the same as the spectral efficiency in bit/s/Hz for nonlinear channels, because of spectral broadening. Another difficulty is the lack of a universal bandwidth definition.

Calculating or approximating capacity

The channel capacity as defined in the previous section is sometimes called the *operational channel capacity*, because it connects to the operation of a communication system. It does however not admit a numerical evaluation, because the maximization over all transmitter and receiver options is prohibitively complex.

For evaluation and analysis, the celebrated *channel coding theorem* can be invoked¹. It states that the operational capacity of any discrete-time channel is equal to the *information channel capacity*⁴

$$C_{dt} = \lim_{N \rightarrow \infty} \frac{1}{N} \sup_{p_{\mathbf{X}}} I(\mathbf{X}; \mathbf{Y}) \quad [\text{bit/channel use}], \quad (1)$$

where

$$I(\mathbf{X}; \mathbf{Y}) = \int p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) \log_2 \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})} d\mathbf{x}d\mathbf{y} \quad (2)$$

is the *mutual information* in bit/block between blocks of N input and output samples \mathbf{X} and \mathbf{Y} , resp., and p denotes the (joint or conditional) distributions of the corresponding blocks. The channel is here represented by a sequence of conditional distributions $p_{\mathbf{Y}|\mathbf{X}}$ for increasing N and the input distribution, which in practice is usually controlled by the modulation format, is similarly represented by a sequence of distributions $p_{\mathbf{X}}$.

The channel coding theorem is an indispens-

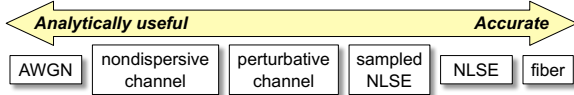


Fig. 1: The dilemma of channel modeling. Available optical channel models are either tractable for capacity analysis (left) or accurate representations of fiber propagation (right), but not both.

able utility in capacity analysis, as it predicts the performance using an ideal (so-called “capacity-achieving”) FEC code without implementing any FEC. There is however no analogous channel coding theorem for continuous-time channels, which is why the vast majority of information-theoretic research has considered discrete-time channels only.

Although the channel coding theorem is extremely powerful, the numerical evaluation of (1) is often beyond reach for complexity reasons. This is because $I(\mathbf{X}; \mathbf{Y})$ includes a multidimensional integral, which often cannot be solved analytically, and furthermore, $p_{\mathbf{Y}}$ in the integrand is itself given by another multidimensional integral given $p_{\mathbf{X}}$ and $p_{\mathbf{Y}|\mathbf{X}}$. Also, $I(\mathbf{X}; \mathbf{Y})$ should be maximized over a multidimensional distribution.

Hence, the channel capacity is unknown for all channels except a few idealized ones. In particular, the capacity of the fiber-optic channel is unknown. In the absence of exact expressions, one can characterize the capacity using estimates, bounds, and asymptotic results.

Channel modeling

As explained in the previous two sections, a mathematical description of the channel is essential for analyzing the channel capacity. In general, the simpler a channel model is, the more accurately can its capacity be characterized, but at the same time, the obtained capacity results may be less relevant for the underlying physical channel. In the context of fiber-optic transmission, this dilemma is illustrated in Fig. 1, where the channel models towards the left are more “information-theory-friendly”⁵ and those towards the right are more realistic. The two extremes are the linear additive white Gaussian noise (AWGN) channel, whose capacity is exactly known, and the actual physical channel, which can only be appraised in experiments. The illustrated classes of models are described, with references, in⁵.

Time discretization

Although all physical communication channels, including the fiber channel, operate in continuous time, the vast majority of information-theoretic

theorems and results apply in discrete time. For example, all channels in Fig. 1 except the two right-most are discrete-time models. Hence, an important and often overlooked element in optical capacity analysis is the discretization of continuous-time channel models. In linear communication systems, which include most wireless and (electric) wired systems, this is not an issue, since the sampling theorem tells us that a strictly bandlimited waveform can be perfectly reproduced from a sequence of its samples. In information-theoretic terms, the discretization process is lossless for linear systems and the continuous- and discrete-time versions of the same channel have the same capacity.

This is not the case for the nonlinear fiber channel. To illustrate this, we simulated propagation of a strictly bandlimited waveform over an optical fiber.* The transmitted and received signals corresponding to 10 transmitted symbols at baudrate $R = 10$ Gbaud are shown in Fig. 2, neglecting noise. They were obtained by split-step Fourier simulations of the discretized NLSE, and the temporal sampling rate R_s was varied. At high enough sampling rate, the simulations converged to the output of the regular (continuous-time) NLSE (thick).

We recall from the sampling theorem that $R_s = R$ is sufficient for linear propagation of complex-valued sinc-shaped signals. This was confirmed in our simulations at low power, when the nonlinearities are insignificant. In the nonlinear example of Fig. 2, the errors introduced by discretization seem, from a visual inspection, to be negligible only when $R_s > 8R$. If $R_s < 4R$, the output loses all resemblance with the ideal (nondiscretized) signal. These observations are somewhat arbitrary and depend significantly on the simulation parameters; the results even change for different transmitted data sequences. No analytic results are yet available about the R_s/R required for negligible discretization errors in the split-step Fourier representation of the NLSE. Therefore, capacity results C_{dt} for the discretized NLSE do not immediately extend to capacities C of the continuous-time NLSE or the true fiber channel.

The need for higher sampling rates R_s in nonlinear transmission simulations can partially be explained by spectral broadening, but no analytic re-

*Parameters: 16-QAM single-wavelength, single-polarization transmission at 10 dBm launch power; raised-cosine pulse shaping with roll-off factor 10 %; 300 km single-mode fiber with ideal distributed amplification; dispersion $\beta_2 = -21.7$ ps²/km; and nonlinearity $\gamma = 1.27$ (W · km)⁻¹.

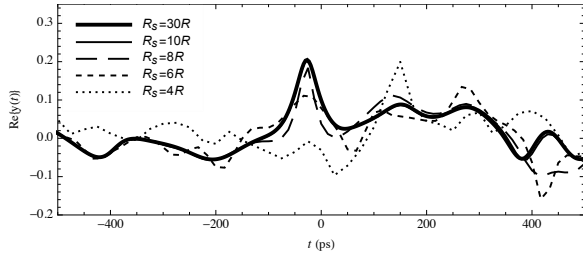


Fig. 2: As the temporal sampling rate in split-step Fourier simulations decreases, the output gets increasingly inaccurate.

lation is yet known between the minimum required R_s and the bandwidth W . For example, the bandwidth of the lightwave whose output is shown in Fig. 2 is $W = 3.2R$.[†] Note that the strict bandwidth is infinite, and hence this bandwidth definition, which is the cornerstone of the sampling theorem, is useless for nonlinear systems.

Lower bounds

A *lower bound* on the capacity can be obtained from the operational capacity definition above or from the information capacity in (1) by relaxing the maximization. In other words, the rate achievable by any (suboptimal) transmitter and receiver is a lower bound, and it can be calculated from the mutual information (2) using any (suboptimal) input distribution.

Another lower-bounding technique, which is often combined with the previous one, is based on *mismatched decoding*. It can be proved that if the channel law $p_{Y|X}$ inside the logarithm in (2) is replaced by an arbitrary conditional distribution, say $q_{Y|X}$, the integral expression cannot increase. Mathematically,

$$I(X; Y) \geq \int p_{X,Y}(x, y) \log_2 \frac{q_{Y|X}(y|x)}{q_Y(y)} dx dy \quad [\text{bit/block}], \quad (3)$$

where $q_Y(y) = \int p_X(x) q_{Y|X}(y|x)$. While (2) is often hard or impossible to compute, as discussed above, (3) may be estimated by Monte Carlo integration, even if $p_{Y|X}$ is intractable. From a practical viewpoint, good mismatched-decoding lower bounds are highly desirable, because those rates are achievable using a receiver optimized for the so-called *auxiliary channel* $q_{Y|X}$. Since this channel is under the system designer's control, it can be chosen to enable a moderate-complexity receiver, whereas the optimal receiver for $p_{Y|X}$, which is needed to achieve

[†]Here W is defined as the smallest frequency range that contains at least 99 % of the signal power at all positions along the fiber.

the mutual information in (2), may be unknown or prohibitively complex.

The most popular auxiliary channel for fiber-optic applications is the Gaussian distribution. This yields the so-called “nonlinear Shannon limit,” which, despite its misleading name, is not a Shannon limit but a lower bound⁶.

Upper bounds

While a lower bound on capacity tells which rate is possible with a selected transmitter–receiver pair, an upper bound states what is impossible using *all* possible such pairs. Therefore, it is generally harder to derive upper bounds.

The most well-known capacity upper bound in optical communications states that the capacity of the discretized NLSE is no greater than $N \log_2(1 + SNR)$ bit/block, where SNR is the launched signal power divided by the total additive noise power, for any R and R_s ⁷, which corresponds to $C \leq R_s \log_2(1 + SNR)$ bit/s or $SE \leq (R_s/W) \log_2(1 + SNR)$ bit/s/Hz. This again illustrates the channel modeling dilemma of Fig. 1: A high R_s implies an accurate model but less interesting bounds on C or SE , and vice versa. If for example $R_s = 8R$ and $W = 3.2R$, as observed in the simulation example above, then $SE \leq 2.5 \log_2(1 + SNR)$ bit/s/Hz. The choices of R_s and W are however somewhat arbitrary, which leaves the interpretation of capacity results in terms of spectral efficiency uncertain.

Acknowledgments

This work was supported in part by the Swedish Research Council (VR no. 2013-5271) and the EPSRC project UNLOC (EP/J017582/1).

References

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3/4, pp. 379–423/623–656, July/Oct. 1948.
- [2] E. Agrell, A. Alvarado, and F. R. Kschischang, “Implications of information theory in optical fibre communications,” *Phil. Trans. R. Soc. A*, vol. 374, no. 2062, p. 20140438, Mar. 2016.
- [3] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, “Capacity limits of optical fiber networks,” *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.
- [5] E. Agrell, G. Durisi, and P. Johannisson, “Information-theory-friendly models for fiber-optic channels: A primer,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2015.
- [6] M. Secondini and E. Forestieri, “Scope and limitations of the nonlinear Shannon limit,” *J. Lightw. Technol.*, vol. 35, no. 4, pp. 893–902, Feb. 2017.
- [7] G. Kramer, M. I. Yousefi, and F. R. Kschischang, “Upper bound on the capacity of a cascade of nonlinear and noisy channels,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2015.