Ripple distribution for nonlinear fiber-optic channels

MARIIA SOROKINA,* STYLIANOS SYGLETOS, AND SERGEI TURITSYN

Aston Institute of Photonic Technologies, Aston University, B4 7ET Birmingham UK
*m.sorokina@aston.ac.uk

Abstract: We demonstrate data rates above the threshold imposed by nonlinearity on conventional optical signals by applying novel probability distribution, which we call ripple distribution, adapted to the properties of the fiber channel. Our results offer a new direction for signal coding, modulation and practical nonlinear distortions compensation algorithms.

OCIS codes: (060.2330) Fiber optics communications; (190.4370) Nonlinear optics, fibers; (110.3055) Information theoretical analysis.

References and links

1. A. Splett, C. Kurtzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in *Tech. Digest of European Conference on Optical Communication* paper MoC2.4. (1993).
2. P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," Nature **411**, 1027–1030 (2001).
3. R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," Phys. Rev. Lett. **101**, 163901 (2008).
4. R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," J. Lightwave Technol. **28**(4), 662–701 (2010).
5. E. E. Narimanov and P. Mitra, "The channel capacity of a fiber optics communication system: perturbation theory," J. Lightwave Technol. **20**(3), 530–537 (2002).
6. D. J. Richardson, "Filling the light pipe," Science **330**, 327–328 (2010).
7. E. Temprana, E. Myslivets, B.P.-P. Kuo, L. Liu, V. Ataie, N. Alic, and S. Radic, "Overcoming Kerr-induced capacity limit in optical fiber transmission," Science **348**, 1445–1448 (2015).
8. P. J. Winzer, "Scaling optical fiber networks: challenges and solutions," Optics and Photonics News, **26**, 28–35 (2015).
9. C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423, 623-656 (1948).
10. P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of non-linear propagation in uncompensated optical transmission links," IEEE Photon. Technol. Lett. **23**(11), 742–744 (2011).
11. I. B. Djordjevic, H. G. Batshon, L. Xu, and T. Wang, "Coded polarization-multiplexed iterative polar modulation (PM-IPM) for beyond 400 Gb/s serial optical transmission," in Proc. Optical Fiber Communication Conference, Los Angeles, CA, Mar. 2010, p. OMK2
12. T. Fehenberger, A. Alvarado Segovia, G. Bocherer, and N. Hanik, "Sensitivity gains by mismatched probabilistic shaping for optical communication systems," IEEE Photon. Technol. Lett. **28**(7) 786–789 (2016).
13. C. Pan and F. R. Kschischang, "Probabilistic 16-QAM Shaping in WDM Systems," in Journal of Lightwave Technology (in press).
14. F. Buchali, G. Bocherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: an experimental demonstration," J. Lightwave Technol. **34**(7), 1599–1609 (2016).
15. E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson "Capacity of a nonlinear optical channel with finite memory," J. Lightwave Technol. **32**(16), 2862–2876 (2014).
16. M. Sorokina, A. Ellis, and S. Turitsyn, "Optical information capacity processing," chapter in All-Optical Signal Processing 325–354 (2015).
17. E. Agrell, G. Durisi, and P. Johannisson, "Information-theory-friendly models for fiberoptic channels: A primer". IEEE Information Theory Workshop (2015).
18. 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, **52**(11), 5008–5022 (2006).
19. H. Song and M. Brandt-Pearce, "A 2-D discrete-time model of physical impairments in wavelength-division multiplexing systems," J. Lightwave Technol. **30**(5), 713–726 (2012).
20. R. Dar, M. Shtaif, and M. Feder, "New bounds on the capacity of the nonlinear fiber-optic channel," Optics Letters **39**, 398–401 (2014).
21. Z. Tao, Y. Zhao, Y. Fan, L. Dou, T. Hoshida, and J. C. Rasmussen "Analytical intrachannel nonlinear models to predict the nonlinear noise waveform," J. Lightwave Technol. **33**(10), 2111–2018 (2015).

22. M. Secondini and E. Forestieri, "Scope and limitations of the nonlinear Shannon limit," online version of 24.10.2016 http://ieeexplore.ieee.org/document/7637002/.

23. M. Secondini, chapter in "Roadmap of optical communications," by E. Agrell et al. Journal of Optics **18** (6), 063002 (2016).

24. M. Sorokina, S. Sygletos, and S. K. Turitsyn, "Shannon capacity of nonlinear communication channels," in Conference on Lasers and Electro-Optics, OSA Technical Digest (2016) (Optical Society of America, 2016), paper SM3F.4.

25. M. Sorokina, S. Sygletos, and S. K. Turitsyn, "Ripple distribution for nonlinear fiber-optic channels," https://arxiv.org/abs/1610.06937

26. P. Johannisson and M. Karlsson, "Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system," J. Lightwave Technol. **31**(8), 1273–1282 (2013).

27. M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," J. Lightwave Technol. **31**(23), 3839–3852 (2013).

28. L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and H. Wymeersch, "A discrete-time model for uncompensated single-channel fiber-optical links," IEEE Trans. Commun. **60**(11), 3440–3450 (2012).

29. R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," Opt. Express **21**(22), 25685–25699 (2013).

30. R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation," J. Lightwave Technol. **33**(5), 1044–1053 (2015).

31. M. Shtaif, R. Dar, A. Mecozzi, and M. Feder, "Nonlinear interference noise in WDM systems and approaches for its cancelation," in Optical Communication (ECOC 2014), 39th European Conference and Exhibition on Optical Communications, paper We1.3.1.

32. 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. Letters **91**, 203901–203904 (2003).

33. M. Secondini, E. Forestieri, and C. R. Menyuk "A combined regular-logarithmic perturbation method for signal-noise interaction in amplified optical systems", J. Lightwave Technol. **27**(16), 3358 – 3369 (2009).

34. K. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers", J. Lightwave Technol. **15**(12), 2232–2241 (1997).

35. M. Schetzen, "The Volterra and Wiener theories of nonlinear systems". Malabar, Florida:Krieger, 2006.

36. A. Amari, P. Ciblat, and Y. Jaouen, "Fifth-order Volterra series based nonlinear equalizer for long-haul high data rate optical fiber communications," Asilomar Conference ACSSC (2014).

37. M. Sorokina, S. Sygletos, and S. Turitsyn, "Sparse identification for nonlinear optical communication systems: SINO method," Opt. Express **24**(26), 30433–30443 (2016).

38. Z. Li, W.-R. Peng, F. Zhu, and Y. Bai, "Optimum quantization of perturbation coefficients for perturbative fiber nonlinearity mitigation," in *Tech. Digest of European Conference on Optical Communication* paper We.1.3.4. (2014).

39. A. Ghazisaeidi and R.-J. Essiambre, "Calculation of coefficients of perturbative nonlinear pre-compensation for Nyquist pulses," in *Tech. Digest of European Conference on Optical Communication paper* We.1.3.3. (2014).

40. S. Verdu and T. S. Han, "A general formula for channel capacity," âĺ IEEE Trans. Inf. Theory **40**(4) 1147âĂŞ-1157 (1994).

41. B. Picinbono, "Second-order complex random vectors and normal distributions," IEEE Transactions on Signal Processing **44** 2637–2640 (1996).

42. E. Agrell and M. Karlsson, "Satellite constellations: towards the nonlinear channel capacity," 25th IEEE Photonics Conference pp. 316–317 (2012).

43. R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, July 2014.

44. D. Rafique and A. Ellis, "Impact of signal-ASE four-wave mixing on the effectiveness of digital back-propagation in 112 Gb/s PM-QPSK systems," Opt. Express **19**(4), 3449–3454 (2011).

45. M. A. Sorokina and S. K. Turitsyn, "Regeneration limit of classical Shannon capacity," Nat. Commun. 5, 3861 (2014).

46. T. Freckmann, R. J. Essiambre, P. J. Winzer, G. J. Foschini, and G. Kramer, "Fiber capacity limits with optimized ring constellations," IEEE Photonics Technology Letters **21**(20), 1496–1498 (2009).

---

1. Introduction

Fiber-optic transmission is the back-bone of global communication systems, yet its capacity, i.e. the maximum error-free transmission rate, remains unknown due to intricate non-linear memory effects [1–8]. In his seminal work in 1948 [9], Shannon proved that Gaussian input signal distribution gives the maximum rate at which information can be transmitted error-free in a linear channel with additive white Gaussian noise, forming the basis for signal coding and modulation in modern communications. The Gaussian distribution for input signal has been widely applied also in nonlinear fiber-optic systems [1–3, 10] creating a number of new coding schemes and modulation formats [11] that could improve the maximum throughput by 0.2 bits compared

to uncoded transmission [12–14]. Yet, Gaussian distribution has shown to be a non-optimum solution for fiber transmission systems, leading to a non-linear threshold at high signal powers.

In fact, it as been pointed out in [15] that Gaussian distribution provides overly pessimistic data rate estimates, below the nonlinear threshold [6–8, 16, 17]. This is because the derived models assume averaging of signal dynamics and lose information about inter-symbol interference effects. Other widely used practical approaches, the so-called "perturbative models with deterministic nonlinearity" [18–21], achieve a first-order perturbative solution of the nonlinear Schrödinger equation by taking into account only the signal-signal interactions in fiber transmission. Such type of nonlinear signal distortion is deterministic, and in principle it can be compensated with some elaborate technical efforts [7], while improving achievable rate [15, 22, 23], making signal-noise interactions the next frontier. Thus, the principal challenge is to provide an accurate analysis of the signal-noise interactions with signal-dependent statistics.

In this paper, we develop a discrete-time finite-memory channel model that takes into account both signal-signal and signal-noise interactions and introduce a novel type of input signal distribution, which we call ripple distribution. These allow us to show that monotonically increasing data rates above the conventional limits can be achieved even in the highly nonlinear regime (i.e. at high signal powers) in the presence of signal-noise interference (expanding results briefly presented in [24, 25]). The findings are also in sharp contrast to estimates based on the Gaussian distribution of input signal and break the notion of "capacity vanishing to zero at high signal power", thus, establishing a new direction for coding and practical nonlinear distortions compensation algorithms.

2.   Finite memory discrete-time channel model.

A typical communication system is presented in Fig. 1(a). At the transmitter, the message is modulated to a discrete time set of constellation symbols (here, 64-QAM plotted in Fig. 1(b) and after pulse shaping it is mapped to a continuous time signal, see Fig. 1(c), which is subsequently launched to a multi-span optical fiber link. The propagation of the continuous-time signal $E(t, z)$ in the optical fiber, see Fig. 1(d), is governed by the well known nonlinear Schrödinger equation (NLSE) (we follow here notations and assumptions used in [3, 4] and refer for details to that papers):

$$\frac{\partial E}{\partial z} = -\frac{\alpha}{2}E - i\frac{\beta_2}{2}\frac{\partial^2 E}{\partial^2 t} + i\gamma|E|^2E + \eta(t, z), \tag{1}$$

where the deterministic distortions, see Fig. 1(e), are introduced by fiber loss α, second-order dispersion parameter β_2, and by Kerr nonlinearity characterized by the coefficient γ. The stochastic distortions are described by the zero-mean additive white Gaussian noise (AWGN) $\eta(t, z)$ of variance $\langle\eta(z, t), \eta^*(z', t')\rangle = \frac{N_D}{L}\delta(z - z')\delta(t - t')$, with N_D and L being the noise spectral density and the transmission length, respectively.

A discrete-time channel model is crucial for an information-theory based analysis as it enables optimization of the mutual information functional $I(\mathbf{X}, \mathbf{Y})$ for deriving the optimum signal distribution $p(\mathbf{X})$, see Fig. 1(f), and for calculating the maximum reliable transmission rate, which is the channel capacity $C = \max I(\mathbf{X}, \mathbf{Y})$. The transition from continuous-time modeling, given by Eq. (1), to discrete-time modelling is not straightforward, since it requires expansion over a complete orthogonal set of basis functions $\{f_k(t)\}$. This is equivalent to matched filter demodulation at the receiver for generating observable discrete-time variables $\{Y_k\}$, see Fig. 1(g). At the transmitter, signal expansion over the carrier pulses is considered, that is $E(t, 0) = \sum_{k=-\infty}^{\infty} X_k f(t - kT)$, where X_k are the complex modulated symbols, $f(t)$ is the time-varying pulse waveform and T is the symbol period. At the receiver, the signal undergoes matched filtering, dispersion compensation (denoted by operator $\mathrm{D}[E]$) and sampling at $t = kT$: $Y_k(\xi) = S^{-1/2}\int dt\mathrm{D}[E(t, \xi)]f(t - kT)$ (with dimensionless coordinate $\xi = z/L_d$ and L_d denoting the dispersion length, while S is signal power), which allows the following
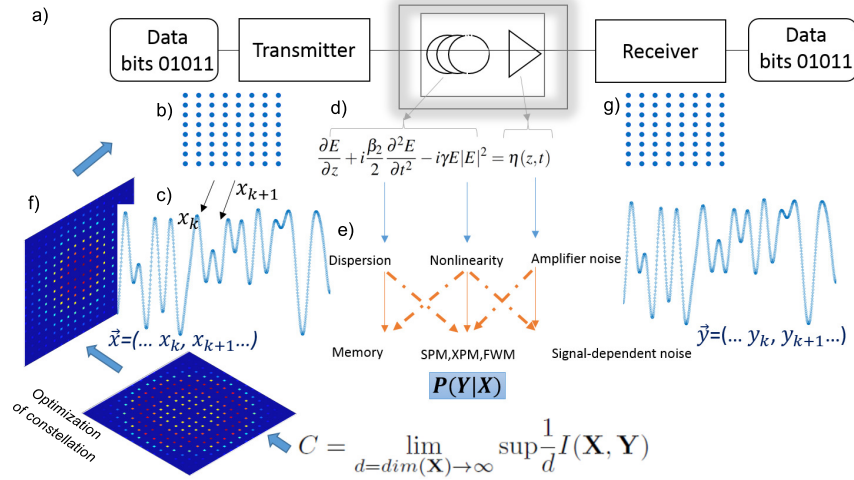
Fig. 1. **Fiber-optic communication system.** a) The fundamental building blocks of a communication system where data is coded to a discrete set of symbols X_k (panel b) and transformed into continuous time form $E(t,0)$ (panel c) to be transmitted via the fiber channel. During transmission the signal is governed by the NLSE (panel d), which results in distortions: dispersion, nonlinearity, noise (panel e), which are reflected in the conditional pdf $P(\mathbf{Y}|\mathbf{X})$, which we can use to optimize constellation and coding (panel f), to receive the maximum achievable transmission rate – channel capacity C. The received signal $E(t,L)$ is processed and sampled Y_k and, finally, decoded to receive the data (panel g).

discrete-time representation of NLSE - multivariate channel model (MCM):

$$Y'_k = \frac{L_d \epsilon}{L} \eta_k + \varepsilon V[Y]_k, \tag{2}$$

$$V[Y]_k = \Psi_s(\xi) \sum_{m,n=-M}^{M} Y_{k+m}(\xi) Y_{k+n}(\xi) Y^*_{k+m+n}(\xi) \tilde{C}_{mn}(\xi)$$

where the AWGN noise term η_k is characterized by the correlation $\langle \eta_k(\xi), \eta^*_{k'}(\xi') \rangle = \Psi_n(\xi) \delta(\xi - \xi') \delta_{kk'}$, where $\Psi_n(\xi) = \lfloor \xi/L_s \rfloor^2 e^{-\alpha mod(\xi, L_s/Ld)}$ is the noise power profile and $\lfloor x \rfloor$ denotes floor function over variable x and $mod(x)$ denotes modulo operation (remainder after division). Here we employ multiple-scale analysis over the two small parameters that characterize the main signal degradation effects, that is: a) the nonlinearity $\varepsilon = L_d/L_{NL}$ (with dispersion length $L_d = T_0^2/\beta_2$ and nonlinearity length $L_{NL} = 1/(\gamma S)$); and b) the noise $\epsilon = \sqrt{N/S} = 1/\sqrt{SNR}$ (which is the reversed square root of the signal-to-noise-ratio (SNR) in the corresponding linear system) with noise power $N = N_D B$ and B denoting the signal bandwidth. Also $\Psi_s(\xi) = e^{-\alpha mod(\xi, L_s/L_d)}$ represents the signal power profile and L_s is the span length. The coupling coefficients $\tilde{C}_{mn}$ define the memory behavior (within a memory window M) of the transmission channel and depend on its physical properties and the signal pulseshape:

$$\tilde{C}_{mn} = i \int \int \int d\omega d\omega_1 d\omega_2 e^{-i\omega_1\omega_2\beta_2 L_d\xi - i\omega_1 mT - i\omega_2 nT} \times \tag{3}$$

$$f^*(\omega)f(\omega_1 + \omega)f(\omega_2 + \omega)f^*(\omega_1 + \omega_2 + \omega)$$

Table 1. Review of previous channel models

model and estimations on capacity	relevance to MCM
signal-signal interactions	
infinite memory Gaussian noise [1, 2, 26, 27] – 1	$Y_k = X_k + \xi_k + \zeta_k,$ $\langle \xi_k \xi_k^* \rangle = \varepsilon^2 \sum_{m,n \neq 0} |C_{mn}|^2$
infinite memory Gaussian noise model [28] – 2	$Y_k = X_k \Phi + \zeta_k$ $\Phi = 1 + \sum_{m,n} |C_{mn}|^2 \varepsilon^2$
finite memory Gaussian noise [15] – 3 non-decreasing capacity	$Y_k = X_k + \xi_k + \zeta_k$ $\langle \xi_k \xi_k^* \rangle = 2\varepsilon^2 \sum_{m,n \neq 0} |C_{mn}|^2 (2M+1)^{-1} \times$ $\sum_{i=k-M}^{k+M} |X_i|^2$
Nonlinear interference noise (NLIN) [29, 30]	1st-order representation of MCM
NLIN with Gaussian approximation [20] which results in concave capacity lower bound	$Y_k = X_k e^{i\theta_k} + \zeta_k$ θ_k, ζ_k - Gaussian variables
NLIN for phase noise [31]	$Y_k = X_k + \sum_m^M X_{k+m} C_m + \zeta_k$ $C_m = \varepsilon \lim_{M \to \infty} \sum_{n=-M}^{M} \langle |X|^2 \rangle C_{mn}$
nonlinear phase noise – 1 [5, 32] capacity equals to linear Shannon limit	$Y_k = X_k + \varepsilon X_k |X_k|^2 C_{00} + \zeta_k$ for small nonlinearity $Y_k = X_k e^{i\gamma L |X_k|^2} + \zeta_k$
signal-noise interactions	
nonlinear phase noise – 2 [32] capacity is monotonically increasing [32]	$(A_k)'_z = \varepsilon C_{00} A_k |A_k|^2 + \eta_k,$ memoryless model
nonlinear signal-noise [33] infinite memory approximation	continuous-time model pdf is derived in continuous-time approximation

The proposed channel model generalizes a number of previous results, see Table 1. Namely, the class of infinite-memory Gaussian noise models, reported in [1, 3, 10, 26], can be received from Eq. (2) after averaging the nonlinear interference term $\langle |C_{mn}|^2 \rangle S^3$, whereas the finite memory model of [15] can be received by averaging of the coupling matrix $\langle |C_{mn}|^2 \rangle$ while keeping the information about interfering symbols X. The model is also related to expansion in Volterra series [34, 35] including multiple order terms [36]. However, the Volterra approach allows to receive continuous time-form, while here the focus is on the discrete-time representation suitable for information theory analysis. Meanwhile, the first order approximation of Eq. (2) was derived in [29] taking into account only the signal-signal terms and ignoring signal-noise interactions. Also, for estimating the capacity the model assumed Gaussian noise statistics, thus losing the exact knowledge of the inter-symbol interference and leading to a concave lower capacity bound that was vanishing to zero at high powers. As it has been shown in [15], such approximations inevitably lead to underestimated capacity bounds. Finally, in [33] the conditional pdf for signal-dependent noise under continuous-time approximation was derived. However, this can not be used for information theory analysis and capacity estimations, as it requires discrete-time treatment and accurate description of the intersymbol interference effects. Therefore, our model is the first finite-memory discrete-time approach that: a) captures pulse-shape and format dependence; b) includes signal-signal and signal-noise interactions; and c) enables incorporation of higher-order terms and derivation of conditional pdfs, which are essential for accurate capacity estimations.

In the main order we have a linear channel with the AWGN noise term ζ_k characterized by the correlation $\langle \zeta_k, \zeta_m^* \rangle = \delta_{km}$. The discrete-time perturbative multivariate channel model has a
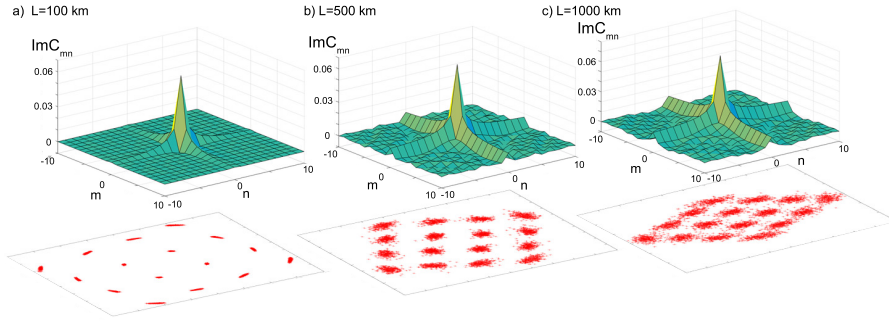
Fig. 2. **Coupling matrix**. For various transmission distances a) $L = 100\ km$, b) $L = 500\ km$, c) $L = 1000\ km$ and the corresponding received constellation diagrams below (for input power 6dBm and span length 100km). We see that the coupling matrix reflects the strength of the inter symbol interaction and its effect on the signal distortion.

tensor form:

$$\mathbf{Y} = \widehat{\mathbf{X}} + \mathbf{M}\zeta + \mathbf{L}\zeta^*, \quad \zeta = \int d\xi \eta(\xi) \tag{4}$$

where the output signal in a vector form $\mathbf{Y} = ...Y_{k-1}\ Y_k\ Y_{k+1}...$ is represented via a deterministic part (signal-signal interactions) $\widehat{\mathbf{X}}$ and a stochastic part (signal-noise interactions) of two nonlinear signal dependent matrices $\mathbf{M}(\widehat{\mathbf{X}})$ and $\mathbf{L}(\widehat{\mathbf{X}})$, respectively, that capture the mixing of signal with noise.

The channel model in Eq. (4) can be rewritten as follows

$$Y_k = \widehat{X}_k + \sum_m M_{k,m}\zeta_m + L_{k,m}\zeta_m^* \tag{5}$$

The first term describes the deterministic Kerr-effect induced inter-symbol interference on the transmitted symbol X_k in the k-th time slot. It can be calculated by solving the deterministic part of Eq. (2) and it can be compensated at the transmitter or the receiver (e.g. the numerical algorithm based on Eq. (2) has been implemented in [37] demonstrating high efficiency and lower complexity compared to the DBP algorithm).

By expanding over small parameter ε we receive:

$$\widehat{X}_k = Y_k^{(0)} + \sum_{N_o=1}^{\infty} \varepsilon^{N_o} Y_k^{(N_o)}, \quad Y_k^{(0)} = X_k \tag{6}$$

The coupling matrix $C_{mn} = \int dz \Psi_s(z) \tilde{C}_{mn}(z)$ governs the signal-signal interactions, particularly those which are responsible for the non-circular distribution of the distortion. This is achieved by taking also into account the pulse-shape impact. The elements of the coupling matrix represent weights of the interference between symbols in different time slots. To demonstrate the effect we considered the transmission of a single-channel in a dispersion unmanaged fiber link. For simplicity, we used Gaussian pulses of 10 ps duration, at full width at half maximum, and a baudrate of 28 GBaud. The link parameters were $\alpha = 0.2dB/km$, $\beta_2 = -20ps^2/km$, $\gamma = 1.3\ 1/W/km$, $L_s = 100km$. The multi-span modeling of the 1st order approximation was verified numerically and experimentally [38, 39]. Fig. 2 shows that the strength of interference between neighboring symbols decays exponentially with their distance (denoted by m, n), whereas the slope is defined by the parameters of the transmission

system. At small transmission distances this interference causes dominance of phase distortion, shown in Fig. 2(a). As the distance increases, the number of interacting symbols also increases and the symbol distortion becomes more circular, see Fig. 2(b), (c). The coupling matrix accurately identifies the interacting symbols and the non-uniform strength of their interference. This is an extremely important characteristic, since by preserving all relevant information about the signal interaction the effect can be completely compensated at the transceiver, and it cannot be achieved with a conventional approach that is based on signal statistics averaging [1, 3, 10, 26].

In Fig. 3(a), we calculate the achievable data rates on a nonlinear transmission system of 1000 km length (without any type of nonlinear compensation), using the existing GN-model [10] and we compare them with our approach based on calculating the variance of nonlinear distortions in Eq. (5) $N_{S-S} = \langle |X - \tilde{X}e^{-i\phi_{nl}}|^2 \rangle = 2\varepsilon^2 \sum_{m,n\neq0} |C_{mn}|^2$ (the limit on summation reflects compensation of the stationary phase shift $\phi_{nl} = 2\varepsilon^2 \sum_m |C_{m0}|^2$). We can see that our model converges to a Gaussian noise model under an infinite-memory approximation, i.e. Eq. (6) is considered as "nonlinear noise" with variance given by N_{S-S} [10].

Deterministic distortions can be compensated with traditional digital back propagation or pre-distortion methods, or alternatively, with the perturbation approach of our developed channel model. Expansion over the parameter ε allows to define the deterministic signal distortion $Y^{(N_o)}$ of order N_o by the recurrence relation:

$$Y_k^{(N_o)} = \sum_{\substack{i,j,l=0 \\ i+j+l=N_o-1}}^{N_o-1} \sum_{m,n} C_{mn} Y_{k+n}^{(i)} Y_{k+m}^{(j)} (Y_{k+n+m}^{(l)})^* \tag{7}$$

The discrete-time characteristic of the approach makes possible the compensation of the nonlinear interference at the receiver without increasing the signal bandwidth. This is in contrast to the traditional pre-distortion techniques that are based on continuous-time waveform processing. After removing the deterministic nonlinear distortions, signal-noise interference becomes the main limitation for increasing the transmitted information rates. Our proposed approach represents the first accurate discrete-time channel model with memory, which can uniquely capture such signal-noise beating effects.

In the main order the matrices are $M_{k,m} = \delta_{k,m}$ and $L_{k,m} = 0$ and we receive a linear AWGN channel; in the next order over parameter ε the signal-noise mixing effects are taken into account

$$M_{k,m} = \epsilon\delta_{k,m} + \epsilon\varepsilon \sum_n K_{n,m-k}(\widehat{X}_{k+n}\widehat{X}_{m+n}^* + \widehat{X}_{m+n}\widehat{X}_{k+n}^*), \tag{8a}$$

$$L_{k,m} = \epsilon\varepsilon \sum_n K_{n,m-k-n}\widehat{X}_{k+n}\widehat{X}_{m-n} \tag{8b}$$

with the matrix

$$K_{mn} = \int dz \sqrt{\Psi_n(z)\Psi_s(z)}\tilde{C}_{mn}(z) \tag{9}$$

To summarize, the model can be applied by the following steps: (i) we calculate the coupling matrix $\tilde{C}_{m,n}(z)$ along the transmission path using Eq. (3), then (ii) we calculate matrices C_{mn} and K_{mn} from expressions below Eq. (2) and Eq. (9); (iii) finally the deterministic distortions are defined for different perturbation orders by setting a zero perturbation order as input signal $Y_k^{(0)} = X_k$ and we calculate the next nonlinearity orders $Y_k^{(1)}, Y_k^{(2)}, ...$ from Eq. (7) and (iv) we finally combine them to obtain the distorted signal $\tilde{X}_k$ using Eq. (6), while (v) the signal-noise interactions are governed by the signal-dependent matrices M_{km} and L_{km} given by Eq. (8). This process allows to characterize signal-signal and signal-noise interactions by a single discrete-time finite memory channel model expressed by Eq. (5) or in tensor form by Eq. (4).

Moreover, the matrix model of Eq. (4) represents a general form that can be easily expanded to cover multi-wavelength operation. In that case, Eq. (5) can be rewritten as follows:

$$Y_{k\theta} = \widehat{X}_{k\theta} + \sum_{m,\alpha} M_{k\theta m\alpha} \zeta_{m\alpha} + L_{k\theta m\alpha} \zeta_{m\alpha}^* \tag{10}$$

here Greek and Latin letters denote frequency and time indexes respectively.

3. Capacity lower bounds

3.1. Capacity and conditional pdf.

To derive the capacity it is necessary to optimize the mutual information functional [40]:

$$C = \lim_{d=dim(\mathbf{X})\to\infty} \sup \frac{1}{d} I(\mathbf{X}, \mathbf{Y}) \tag{11}$$

under the power constraint: $\int d\mathbf{x} ||\mathbf{x}||^2 P_{\mathbf{x}} = d$. Here input and output vectors of length d are given by: $\mathbf{X} = (X_1, ..., X_d)$ and similarly for $\mathbf{Y} = (Y_1, ..., Y_d)$. To optimize the mutual information over input pdf one must calculate the multivariate conditional pdf that takes into account the memory effects defined by the channel model of Eq. (4). Next, we derive the *conditional pdf* for the transmitted symbols. The channel model of Eq. (4) represents a mixing of signal and noise components as a result of the intersymbol interference. A linear combination of univariate independent and identically distributed normal vectors can be represented as a complex normal distribution [41]

$$P(\mathbf{Y}|\mathbf{X}) = (\pi)^{-d} (|\mathbf{\Gamma}||\mathbf{P}|)^{-1/2} \tag{12}$$

$$\exp\left[-\frac{1}{2}[(\mathbf{y}-\widehat{\mathbf{x}})^H, (\mathbf{y}-\widehat{\mathbf{x}})^T] \begin{pmatrix} \mathbf{\Gamma} & \mathbf{\Upsilon} \\ \mathbf{\Upsilon}^H & \mathbf{\Gamma}^* \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}-\widehat{\mathbf{x}} \\ \mathbf{y}^*-\widehat{\mathbf{x}}^* \end{pmatrix} \right]$$

Notations: * means the complex conjugate, ^{T} means transposition, and ^{H} means transposition and complex conjugate.

The covariance matrix $\mathbf{\Gamma}$ (real, symmetric, non-negative definite and Hermitian) and relation matrix $\mathbf{\Upsilon}$ (real and symmetric) are given as

$$\mathbf{\Gamma} = E[(\mathbf{Y}-\widehat{\mathbf{X}})(\mathbf{Y}-\widehat{\mathbf{X}})^H], \quad \mathbf{\Upsilon} = E[(\mathbf{Y}-\widehat{\mathbf{X}})(\mathbf{Y}-\widehat{\mathbf{X}})^T] \tag{13}$$

$$\mathbf{P} = \mathbf{\Gamma}^* - \mathbf{\Upsilon}^H \mathbf{\Gamma}^{-1} \mathbf{\Upsilon}$$

Recalling Eq. (4)

$$\mathbf{\Gamma} = (\mathbf{M}\mathbf{M}^H + \mathbf{L}\mathbf{L}^H), \quad \mathbf{\Upsilon} = (\mathbf{M}\mathbf{L}^T + \mathbf{L}\mathbf{M}^T) \tag{14}$$

This is the first result that presents the conditional pdf derived from a discrete-time model with memory that has been accurately defined by the nonlinear properties of the fiber optic channel. This approach allows to capture any non-circular behavior of signal distortions and it contains precise information about intersymbol and signal-noise interfering effect. In the particular case of $\mathbf{\Upsilon} = 0$ and $E(\mathbf{Y}) = 0$ we have circularly symmetric complex normal distribution $\mathbf{Y} \sim C\mathcal{N}(0, \mathbf{\Gamma})$, i.e. $P(\mathbf{y}) = (\pi)^{-d}|\mathbf{\Gamma}|^{-1} e^{-\mathbf{y}^H \mathbf{\Gamma}^{-1} \mathbf{y}}$.

3.2. Ripple distribution and monotonically increasing lower bound.

Once the conditional pdf is determined, one can use different input pdfs to determine lower bounds on capacity. Here we introduce ripple distribution as the input signal distribution. To construct this distribution we consider a set of $\alpha = 1..q$ Gaussian distributions with uniform
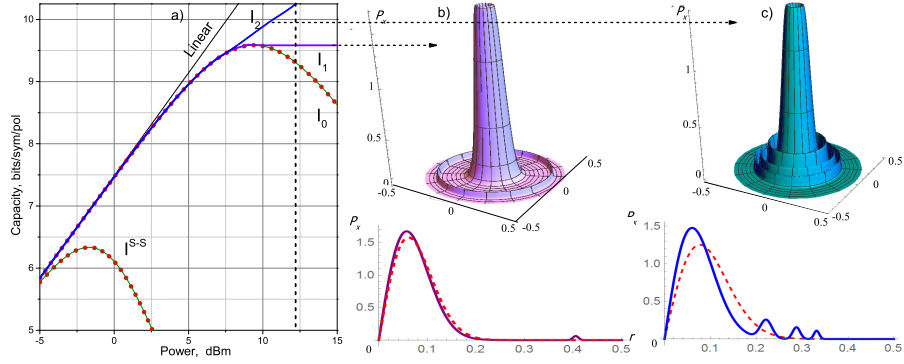
Fig. 3. **Capacity and ripple distribution**. a) Capacity lower bounds for uncompensated signal-signal (S-S) distortions I^{S-S} (GN model (red, dotted) and the proposed model (green)) and compensated deterministic distortions with account of signal-noise (S-N) interference. Previous lower bound (denoted by I_0) decreases to zero, whereas ripple distribution as input pdf (Eq. (15)) allows to achieve higher monotonically increasing bounds, denoted by I_1 and I_2 for the increased number of ripples in an input pdf as shown in panels b) and c) correspondingly (Gaussian pdf is shown in red dashed lines for comparison).

Table2: Statistical properties of ripple distribution

	Ripple	Normal
pdf	$\frac{r}{\pi\sigma}e^{-\frac{r^2+\rho^2}{\sigma}}I_0\left(\frac{2r\rho}{\sigma}\right)$	$\frac{1}{\pi\sigma}\exp\left(-\frac{(x-x_0)^2+(y-y_0)^2}{\sigma}\right)$
mean	0	$\rho = \sqrt{x_0^2 + y_0^2}$
mode	ρ	$\rho = \sqrt{x_0^2 + y_0^2}$
2-nd order non-central moment	$\rho^2 + \sigma$	$\rho^2 + \sigma$
4-th order non-central moment	$2\sigma^2 + 4\sigma\rho^2 + \rho^4$	$3\sigma^2 + 6\rho^2\sigma + \rho^4$
Kurtosis	$1 + \sigma\frac{\sigma+2\rho^2}{(\sigma+\rho^2)^2}$	$1 + 2\sigma\frac{\sigma+2\rho^2}{(\sigma+\rho^2)^2}$

phase, each of which is localized around a different amplitude level (mode) ρ_α and it has different variance σ_α and weight p_α, so that in polar coordinates it is represented as:

$$P_\mathbf{X} = \prod_{i=1}^{d} P_{X_i}, \quad P_{X_i}(x_i = \{r_i, \varphi_i\}) = \sum_{\alpha=1}^{q} \frac{r_i p_\alpha}{\pi\sigma_\alpha} e^{-\frac{r_i^2+\rho_\alpha^2}{S_\alpha}} I_0\left(\frac{2r_i\rho_\alpha}{\sigma_\alpha}\right) \tag{15}$$

here the Latin alphabet is used to denote the time-index, and the Greek, the coding level. Note, that random variables are denoted by uppercase letters X whereas their (deterministic) outcomes by the lowercase letter x.

Note that the ripple distribution is a new kind of probability distribution. Although different shaping algorithms have been previously used, they were mostly focusing on applying Gaussian pdfs to discrete [12–14] or quasi-continuous (e.g. rings [4]) formats. As a result, all of them have led to capacity lower bounds vanishing to zero at high powers. In [15] a different distribution with a heavy tale has been used (similar to "Satellite constellations" [42]) which enabled to receive a lower capacity bound saturating to plateau. A similar plateau-like behavior for signal-noise case has been claimed in [23], however the distribution has not been reported. It has been shown that Gaussian distribution is not the optimal input distribution in the nonlinear optical channel and distributions with heavier tails might be more beneficial [15, 43]. The statistical properties of the

ripple distribution with one ripple are presented in the Table 2.

In the simplest case of $q = 1$, with mean $\rho_1 = 0$ and variance σ_1 we receive the conventional result:

$$P_{\mathbf{X}} = \prod_{i=1}^{d} \frac{r_i}{\pi \sigma_1} e^{-\frac{r_i^2}{\sigma_1}} \tag{16}$$

which is the Gaussian pdf with zero mean and variance σ_1 in polar coordinates. The resulting lower bound is

$$I_0 = \log_2 \left(1 + \frac{\sigma_1}{\epsilon^2 + C_{nl}\sigma_1^2} \right) \tag{17}$$

which is plotted in Fig. 3(a), where $C_{nl} = 6\varepsilon^2 \epsilon^2 \sum_{m,n \neq 0} |K_{mn}|^2$, and is in agreement with the asymptotic closed form expression of [44]. The proposed model is applicable in a regime of operation where spectral broadening is small, thus fulfilling the condition $L_D < L_{NL}$. For the parameters used in Fig.3 and a maximum signal power 13 dBm $L_D/L_{NL} \simeq 1$ numerical simulations have shown that the spectral broadening is indeed small (limit set by the vertical line in Fig. 3(a)). Further increase of signal power above this level will violate the aforementioned condition and the model will need to be further expanded.

For $\epsilon < C_{nl}^{-1/2}$ we consider a two-ripple distribution ($q = 2$): one centered around zero power level and having variance $\sigma_1 = \epsilon C_{nl}^{-1/2}$, whereas the second ripple has a small variance $\sigma_2 \ll \rho^2$ with vanishingly small probability $\delta \to 0$ and centered around distant power level $\rho^2 \gg \sigma_1$ (see Fig. 3(b)). Thus, we consider $p_{\alpha=1,2} = \{1 - \delta, \delta\}$, $\sigma_{\alpha=1,2} = \{\sigma_1, \sigma_2\}$, and $\rho_{\alpha=1,2} = \{0, \rho\}$.

The aforementioned constrains on probability for $\delta \to 0$ can be rewritten as:

$$1 = \int d\mathbf{x} P_{\mathbf{x}} \simeq p_1^d + d p_1^{d-1} p_2$$

$$d = \int d\mathbf{x} ||\mathbf{x}||^2 P_{\mathbf{x}} \simeq d p_1^d \sigma_1 + d p_1^{d-1} p_2 ((d-1)\sigma_1 + \sigma_2 + \rho^2)$$

and the corresponding lower bound on capacity (assuming $\rho^2 \gg \sigma_1$, $\rho^2 \gg \sigma_2$, $\delta \to 0$) is found as

$$I_1 = \lim_{\substack{\delta \to 0 \\ d \to \infty}} p_1^d \log_2 \left(1 + \frac{\sigma_1}{\epsilon^2 + 6\varepsilon^2 \epsilon^2 \sum_{m,n} |K_{mn}|^2 \sigma_1^2} \right) - p_1^d \log_2 p_1 - p_0^{d-1} p_2 \log_2(p_1^{d-1} p_2) + \frac{p_1^{d-1} p_2}{d} \sum_{l=-d/2}^{d/2}$$

$$\left[\log_2 \left(\frac{2\sqrt{\pi \rho^2 \sigma_2}}{\epsilon^2 + 6\varepsilon^2 \epsilon^2 \sum_{m,n \neq l} |K_{mn}|^2 \sigma_1^2 + 12\varepsilon^2 \epsilon^2 \sum_m |K_{ml}|^2 \sigma_1(\sigma_2 + \rho^2) + 6\varepsilon^2 \epsilon^2 |K_{ll}|^2(\sigma_2 + \rho^2)^2} \right) + \right.$$

$$\left. (d-1) \log_2 \left(1 + \frac{\sigma_1}{\epsilon^2 + 6\varepsilon^2 \epsilon^2 \sum_{m,n \neq l} |K_{mn}|^2 \sigma_1^2 + 12\varepsilon^2 \epsilon^2 \sum_m |K_{ml}|^2 \sigma_1(\sigma_2 + \rho^2) + 6\varepsilon^2 \epsilon^2 |K_{ll}|^2(\sigma_2 + \rho^2)^2} \right) \right]$$

given the adopted notation the equation is simplified as

$$I_1 = (1-\delta) \log_2 \left(1 + \frac{1}{2\epsilon^{-1}\sqrt{C_{nl}}} \right) + \delta \log_2 \left(\frac{2\sqrt{\pi \rho^2 \sigma_2}}{\sigma_1} \right) + \delta \log_2 e - \delta \log_2 \delta - \delta(\sigma_2 + \rho^2 - \sigma_1)\epsilon^{-1} C_{nl}^{1/2} \log_2 e$$

optimization over free δ results in

$$I_1 = \log_2 \left(1 + \frac{1}{2\epsilon\sqrt{C_{nl}}} \right) + \sqrt{2\pi \rho^2 \epsilon^{-1} C_{nl}^{1/2}} e^{-\rho^2 \epsilon^{-1} C_{nl}^{1/2}} \log_2 e \tag{18}$$

and parameter ρ can be found from power requirement: $\sqrt{2\pi \rho^2 \epsilon^{-1} C_{nl}^{1/2}} \rho^2 e^{-\rho^2 \epsilon^{-1} C_{nl}^{1/2}} = 1 - \sigma_1$.

The result shows that with the considered suboptimal input pdf, we can receive a monotonically increasing lower capacity bound, which is asymptotically close to the plateau level. Therefore, we proved that signal-noise effects do not degrade the capacity with the increase of the signal power. As we show below, further optimization of the pdf can only improve this bound.

Note that the proposed channel model and the corresponding conditional pdf enable *numerical calculation* of lower bounds (i.e. the mutual information for a given input pdf), without invoking approximations or asymptotics. For a discrete-time finite-channel memory model with signal-signal interactions this methodology was demonstrated in [15] for a single-channel and expanded recently in [23] for multichannel (WDM) operation. Here, we follow the same recipe as in [15], but we also include the signal-noise interactions, resulting in a conditional pdf which captures signal-noise interactions. The calculation is as follows: (i) we derive the discrete-time finite-memory model (Eq. (5)); (ii) derive conditional pdf (Eq. (12)); (iii) we select the input pdf (Eq. (15)); (iv) we calculate the corresponding mutual information (Eq. (11)).

We consider a ripple distribution for larger number of levels q (e.g. Fig. 3(c)) and with adjustable variances and centers. Using Monte Carlo simulations for multidimensional integration in Eq. (11) we have found that the numerical optimization (i.e. selecting the optimum parameters p_α, σ_α, ρ_α and varying the number q for every power S) results in a monotonically increasing I_2 numerical bound, as shown in Fig. 3(a)). We stress that these are the first monotonically increasing lower bound calculated asymptotically I_1 and numerically I_2.

As we have shown, ripple distribution can provide the mutual information that is monotonically increasing with signal power. This distribution has similarities with the optimum input pdf for another type of nonlinear channels - regenerative channels [45].

Note that the ripple distribution can be considered as a generalization of the so-called ring constellations [46] under the condition that the variance of each ripple S_α tends to zero. While ring distribution has been shown to provide improved lower bounds than the Gaussian distribution, these bounds are decreasing with signal power (for uncompensated signal-signal interactions). On the other hand, when taking into account only two ripples and setting the variance of the second ripple to zero one can reproduce the so-called satellite constellations, the optimization of which [42] results in non-decreasing lower bounds that saturate to a plateau level. Contrary to both of these cases, a ripple distribution with finite variance introduces additional degrees of freedom in the optimization of the constellation, which results in increasing lower bounds.

4. Conclusion

We demonstrated that channel capacity is a monotonically increasing function of signal power when ripple distribution of the input signal is used. This was achieved by deriving a novel finite memory discrete-time multivariate channel model capable of describing nonlinear signal-signal and signal-noise inter-symbol interfering effects in transmission links. Both model and ripple distribution form a complete information-theoretic tool for capacity calculations, which can be used for the practical design of efficient compensation algorithms and coding schemes tailored by nonlinearity in fiber optic communication channels.

Funding