New Bounds on the Capacity of the Nonlinear Fiber-Optic Channel

Ronen Dar, Mark Shtaif*, and Meir Feder

School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel
* *shtaif@tauex.tau.ac.il*

We revisit the problem of estimating the nonlinear channel capacity of fiber-optic systems. By taking advantage of the fact that a large fraction of the nonlinear interference between different wavelength-division multiplexed channels manifests itself as phase noise, and by accounting for the long temporal correlations of this noise, we show that the capacity is notably higher than what is currently assumed. This advantage is translated into the doubling of the link distance for a fixed transmission rate.

Estimation of the fiber-optic channel capacity has come to be one of the most challenging and important problems in the field of optical communications [1–7]. Recently, its importance has grown even higher as the latest capacity estimates are being rapidly approached by the rates of commercial communications systems [8]. The difficulty in estimating the capacity of the fiber-optic channel is mostly due to the effect of fiber nonlinearity which generates complicated distortions of the transmitted optical waveforms. Perhaps the most comprehensive and familiar attempt of estimating the fiber-channel capacity to date is due to Essiambre et al. [3], where it was argued that, under plausible assumptions on network architecture, nonlinear interference between different wavelength-division multiplexed (WDM) channels must be treated as noise, which was then identified as the predominant nonlinear factor in limiting the capacity of the fiber-optic channel. This point of view has been adopted by most subsequent studies [4–6], and we also adopt it in the study presented herein.

A common feature of capacity estimates published so far is that they treat the nonlinear noise as additive, white and independent of the data transmitted on the channel of interest. In reality, in the presence of chromatic dispersion, different WDM channels propagate at different velocities so that every symbol in the channel of interest interacts with multiple symbols of every interfering channel. Consequently, adjacent symbols in the channel of interest are disturbed by essentially the same collection of interfering pulses and therefore they are affected by nonlinearity in a highly correlated manner. In addition, as has been recently demonstrated in [6,9], one of the most pronounced manifestations of nonlinearity is in the form of phase noise due to cross-phase modulation (XPM). The dominance of the phase noise nature of the nonlinear interference is particularly pronounced in systems with distributed gain, which is the regime where the capacity of the fiber-optic channel has been evaluated [3], and which is also assumed in the present work.

We demonstrate in what follows that by taking advantage of the long temporal correlations that allow the cancelation of nonlinear phase noise, it is possible to communicate at a higher rate than predicted in [3], or equivalently (almost) double the distance achievable at a given rate of communications. We stress that practical methods for canceling the nonlinear phase noise are not discussed in this paper as we are only interested in the capacity implications. We also note that nonlinear phase noise is canceled inadvertently in coherent optical systems where an appropriately fast phase tracking algorithm is deployed.

We start by expressing the received signal samples after coherent detection and matched filtering as

$$y_j = x_j \exp(i\theta_j) + n_j^{\mathrm{NL}} + n_j, \tag{1}$$

where j is the time index and the term n_j^{NL} accounts for all nonlinear noise contributions that do not manifest themselves as phase noise. As in [3–6] we assume that the samples n_j^{NL} are zero-mean statistically independent complex Gaussian variables with variance σ_{NL}^2. A similar assumption holds for the amplified spontaneous emission (ASE) samples n_j, whose variance is denoted by σ_{ASE}^2. All three noise contributions θ_j, n_j^{NL}, and n_j are assumed to be statistically independent of each other. All of the above assumptions, regarding the whiteness of n_j^{NL} and n_j, the statistical independence of all noise contributions and the Gaussianity of n_j^{NL} constitute a ***worst case*** in terms of the resultant capacity [10, Ch. 10] and hence they are in accord with our goal of deriving a capacity lower bound. Finally, consistently with what is suggested by the analysis in [6], we will also assume that θ_j is a Gaussian distributed variable and its variance will be denoted by σ_θ^2, whose expression can be found in [6,9].

For arriving at an analytical lower bound for the capacity, we assume that the nonlinear phase noise θ_j is blockwise constant. In other words, it is assumed that

the noise θ_j does not change at all within a block of N symbols and then in the subsequent block it changes in a statistically independent manner. The assumption that θ_j does not change within a block is consistent with the very long temporal correlations of the phase noise, as was demonstrated in [9, 11, 12]. The assumption of statistical independence of θ_j in adjacent blocks is again a worst-case scenario which is in accord with our interest in a lower bound.

The capacity of the block-wise independent phase noise channel (1) is given by

$$\mathrm{C} = \frac{1}{N} \sup_{p_{\underline{X}}} \; I(\underline{X};\underline{Y}), \tag{2}$$

where $\underline{X}$ and $\underline{Y}$ are column random vectors representing a block of N channel inputs and outputs, respectively, in which the phase noise is constant. The term $I(\underline{X};\underline{Y}) = h(\underline{Y}) - h(\underline{Y}|\underline{X})$ is the mutual information between the channel's input and output, where $h(\cdot)$ is the differential entropy. The supremum is over all input distributions satisfying the power constraint $\mathbb{E}[\|\underline{X}\|^2] = NP$. In order to obtain a lower bound on the capacity we assign $\underline{X}$ a circular Gaussian distribution with statistically independent elements. Note that with this input distribution, and taking into account the fact that n_j^{NL} and n_j are uncorrelated with the input signal [6,9], our assumption that these quantities are white circular Gaussian constitutes a worst case scenario from the standpoint of the resultant capacity [10, Ch. 10, Ex. 1]. In this case $\underline{Y}$ is also a circularly symmetric complex Gaussian vector with differential entropy $h(\underline{Y}) = N\log_2(\pi e(P+\sigma^2_{\mathrm{eff}}))$, where σ^2_{eff} is the variance of the effective additive noise, i.e., $\sigma^2_{\mathrm{eff}} = \sigma^2_{\mathrm{ASE}} + \sigma^2_{\mathrm{NL}}$. The conditional distribution of $\underline{Y}$ given $\underline{X}$ is obviously not Gaussian (see Eq. (1)), but since the Gaussian distribution maximizes the differential entropy of a vector of zero-mean random variables with a given covariance matrix [10, Ch. 9, Theorem 9.6.5], the differential entropy $h(\underline{Y}|\underline{X})$ satisfies

$$\begin{aligned} h(\underline{Y}|\underline{X}) &= \mathbb{E}_{\underline{x}}\big(h(\underline{Y}|\underline{X}=\underline{x})\big) &(3)\\ &\le \frac{1}{2}\mathbb{E}_{\underline{x}}\big(\log_2\det(2\pi e\, Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}})\big)\;, &(4)\end{aligned}$$

where $\hat{\underline{Y}} = \begin{bmatrix}\Re\mathfrak{e}(\underline{Y})\\ \Im\mathfrak{m}(\underline{Y})\end{bmatrix}$ and $Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}}$ is the covariance matrix of $\hat{\underline{Y}}$ given $\underline{X}=\underline{x}$. By applying some algebraic manipulations the determinant of $Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}}$ can be shown to satisfy

$$\det(Q_{\hat{\underline{Y}}|\underline{X}=\underline{x}}) = (\frac{\sigma^2_{\mathrm{eff}}}{2})^{2N}(1+2\frac{\|\underline{x}\|^2}{\sigma^2_{\mathrm{eff}}}\sigma_c^2)(1+2\frac{\|\underline{x}\|^2}{\sigma^2_{\mathrm{eff}}}\sigma_s^2)\;, \tag{5}$$

where the terms $\sigma_c^2 = 0.5(1-e^{-\sigma_\theta^2})^2$ and $\sigma_s^2 = 0.5(1-e^{-2\sigma_\theta^2})$ are the variances of $\cos(\theta)$ and $\sin(\theta)$, respectively, and their calculation relies on the Gaussianity of θ_j. Note that throughout this paper the Gaussianity assumption of the phase noise is needed only here, for calculating σ_c^2 and σ_s^2. Finally, by plugging (5) into (4), the following capacity lower bound is obtained

$$\begin{aligned} \mathrm{C} \;\ge\; & \log_2\left(1+\frac{P}{\sigma^2_{\mathrm{eff}}}\right)\\ & - \frac{1}{2N}\mathbb{E}_\upsilon\left\{\log_2\left(1+\upsilon\sigma_c^2\frac{P}{\sigma^2_{\mathrm{eff}}}\right)\right\}\\ & - \frac{1}{2N}\mathbb{E}_\upsilon\left\{\log_2\left(1+\upsilon\sigma_s^2\frac{P}{\sigma^2_{\mathrm{eff}}}\right)\right\}, \end{aligned} \tag{6}$$

where the symbol $\mathbb{E}_\upsilon$ stands for ensemble averaging with respect to a standard Chi-square distributed variable υ with $2N$ degrees of freedom. Notice that the first line on the right-hand-side of Eq. (6) follows from treating the nonlinear noise as white circular Gaussian – similarly to the analysis in [3]. Yet, the difference is that in our case σ^2_{NL} corresponds only to the part of the nonlinear noise that does not manifest itself as phase noise and hence it is smaller than the nonlinear noise that is accounted for in [3]. The effect of phase noise on the capacity is captured in our case by the bottom two lines on the right-hand-side of (6). This capacity loss, which may be viewed as a rate reduction needed for tracking the phase noise, vanishes when the phase exhibits very long term correlations (i.e., when $N \to \infty$).

We have performed a set of numerical simulations in order to extract the variances σ^2_θ and σ^2_{eff} and to obtain lower bounds on the capacity of the nonlinear fiber channel. The simulations were performed using the parameters of a standard single mode fiber; dispersion $D = 17$ ps/nm/km, attenuation of 0.2 dB/km, nonlinear coefficient $\gamma = 1.27\ \mathrm{W}^{-1}\mathrm{km}^{-1}$ and signal wavelength $\lambda_0 = 1.55\ \mu$m. Perfectly distributed and quantum limited (i.e. fully inverted) amplification with spontaneous emission factor $n_{sp} = 1$ was assumed. Sinc-shaped pulses with a perfectly square 100 GHz wide spectrum were used for transmission and the spacing between adjacent WDM channels was 102 GHz (i.e leaving a 2 GHz guard band). The number of simulated WDM channels was 5, with the central channel being the channel of interest. All of the above assumptions are identical to those made by Essiambre et al. in the computation of the capacity lower bound reported in [3]. The number of simulated symbols in each run was 8192 for the 500km system and 16384 for the 1000km and 2000km systems. Up to 500 runs (each with independent and random data symbols) were performed with each set of system parameters, so as to accumulate sufficient statistics. We assumed a circularly symmetric complex Gaussian distribution of points in the transmitted constellation. This constellation was used to derive our capacity lower bound (6). At the receiver, the central channel was filtered out with a perfectly square filter (which is also the matched filter with sinc pulses) and back-propagated ideally (using the same step-size criteria as in the forward propagation). Then, the signal was optimally sampled and analyzed. As in [3], all simulations have been performed with the

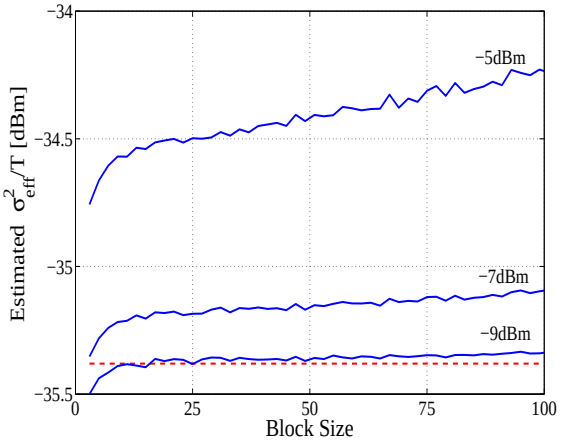

Fig. 1. The numerically estimated $\sigma^2_{\rm eff}$ (normalized by T) vs. block-size in a 500km link for input average power levels of -9dBm, -7dBm and -5dBm. Red dashed line shows $\sigma^2_{\rm ASE}/T$. Due to insufficient statistics for small values of N, the estimated $\sigma^2_{\rm eff}$ grows rapidly with block-size. Then, when the accumulated statistics is sufficient, the growth is much slower and it is due to the fact that phase fluctuations inflate the estimated $\sigma^2_{\rm eff}$.

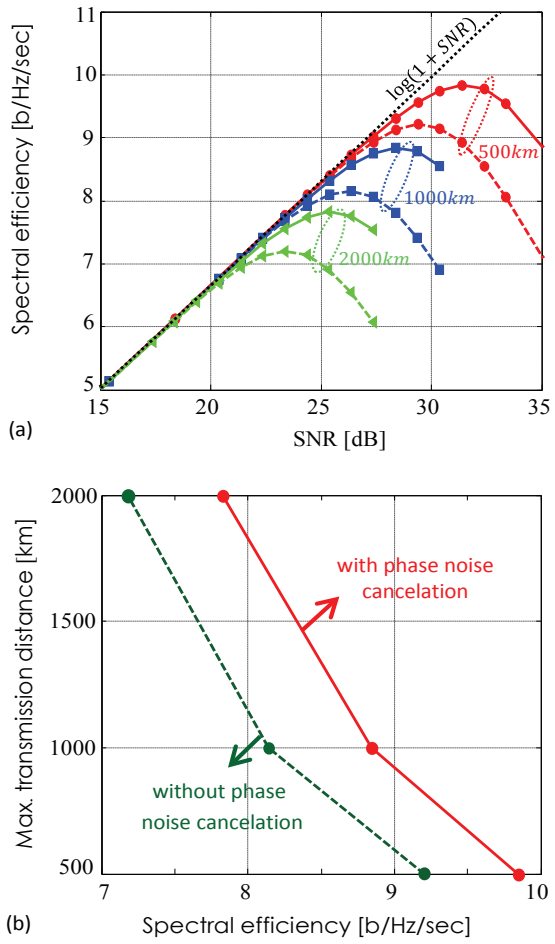

Fig. 2. (a) Capacity lower bound vs. linear SNR for 500km (red dots), 1000km (blue squares), and 2000km (green triangles). Dashed curves result from treating the entire nonlinear noise as noise. Solid curves represent the new bounds derived here. Dotted curve represents the Shannon limit $\log_2(1+\text{SNR})$. (b) The maximum achievable transmission distance as a function of spectral efficiency with (solid) and without (dashed) phase noise cancelation.

scalar nonlinear Schrödinger equation and correspond to a single polarization. Consideration of the polarization multiplexed case, where cross-polarization phenomena are directly accounted for in the simulations and in the analysis, is left for future study.

In order to extract the angle θ we exploit the fact that the nonlinear phase noise changes very slowly on the scale of the symbol duration, and apply the least-squares estimation procedure. Namely, we evaluated $\exp(i\theta)$ by averaging the variable $x_j^* y_j$ over $N = 50$ adjacent symbols and then normalizing the absolute value of the averaged quantity to 1, so as to ensure that we are only extracting phase noise. The estimate of $\exp(i\theta)$ will be denoted by $\exp(i\hat{\theta})$ in what follows. We then subtracted $x_j \exp(i\hat{\theta})$ from y_j to obtain $n_j^{\rm NL} + n_j$ and to evaluate $\sigma^2_{\rm eff}$. Note that the choice of N affects the estimated noise variance in two ways. On the one hand, the estimation of $\sigma^2_{\rm eff}$ improves as N increases (the mean square error of the estimation is proportional to N^{-1}). On the other hand, the assumption of constant phase noise becomes less accurate as N increases. As a result the variations of the phase noise inflate the estimate of $\sigma^2_{\rm eff}$ and reduce the tightness of our capacity lower bound. Fig. 1 shows the dependence of the estimated value of $\sigma^2_{\rm eff}$ on the assumed block-size N for several values of average signal power per-channel. The various curves share an important and very distinct feature. In all cases, the estimated value of $\sigma^2_{\rm eff}$ grows with N at small N values and then it abruptly transitions to a much slower rate of growth. The fast growth in the first stage is due to the lack of sufficient statistics at small N values, whereas the slow growth in the second stage is due to the slow variations of the phase noise whose significance increases with increasing block-size. Our choice of $N = 50$ is always higher than the value of N that corresponds to the transition between the two growth rates, thereby guaranteeing that sufficient statistics is used in all cases (albeit at the expense of a slightly overestimated $\sigma^2_{\rm eff}$). Finally, the variance σ^2_θ was evaluated by extracting $\hat{\theta}_j$ from a sliding window average (of width $N = 50$) performed over all simulated symbols.

Fig. 2a shows the capacity lower bound curves as a function of the linear SNR (which is the ratio between the average signal power, P, and the power of the ASE noise within the channel bandwidth, $\sigma^2_{\rm ASE}$). The dashed curves correspond to the case in which we do not separate the phase noise and treat the entire nonlinear distor-

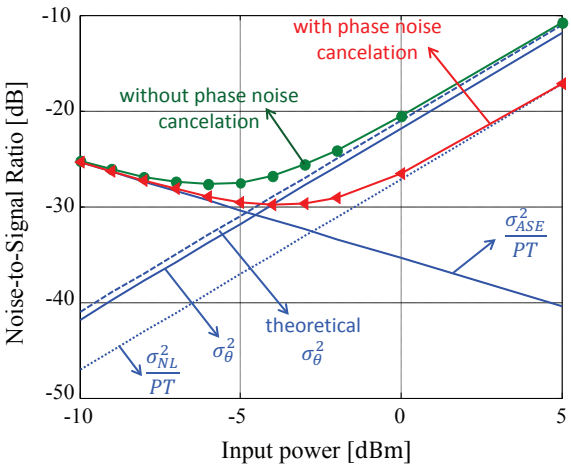

Fig. 3. Noise-to-signal ratio vs. average input power in a 500km link. Decreasing solid line shows $\sigma^2_{\rm ASE}/(PT)$, increasing solid line shows σ^2_θ. Dashed line is the theoretical expression for σ^2_θ found in [6, 9]. Dotted line is $\sigma^2_{\rm NL}/(PT)$. Triangles and dots show $\sigma^2_{\rm eff}/(PT)$ with and without phase noise cancelation, respectively.

tions as noise. These curves are consistent with the study reported in [3], except that our usage of Gaussian modulation leads to slightly higher capacity values. The solid curves represent our new lower bound, achieving a peak capacity that is higher by approximately 0.7 bit/sec/Hz in all cases. In Fig. 2b we invert the peak capacity results of Fig. 2a so as to plot the maximum achievable system length as a function of the spectral efficiency. As is evident from the figure, the achievable system length is approximately doubled by exploiting our scheme. Finally in Fig. 3 we show the various noise contributions as a function of the average power per channel in the case of a 500km link. The monotonically decreasing blue curve shows the noise to signal ratio (NSR) $\sigma^2_{\rm ASE}/(PT)$ due to the ASE noise by itself. The monotonically increasing solid blue curve shows σ^2_θ, which in the limit of small variations in θ_j, represents the NSR due to phase noise. The dashed blue line is the theoretical expression for σ^2_θ as given in [6, 9], which is seen to be in very good agreement with our numerical result. The blue dotted line shows the NSR due to the residual nonlinear noise $\sigma^2_{\rm NL}/(PT)$ after separating the phase noise. As is evident in the figure, separation of nonlinear phase noise reduces the nonlinear noise by approximately 6dB. Triangles and dots show $\sigma^2_{\rm eff}/(PT)$ with and without phase noise cancelation, respectively. Evidently, the minimum effective NSR of the system is improved by approximately 2dB. We note that to facilitate the distinction between the noise contributions the simulation that produced Fig. 3 was performed without ASE propagation. In Figs. 2a and 2b ASE noise was propagated, although similar results were observed when the ASE was added at the end. We also note that these figures were obtained using more statistics than those presented in [11].

To conclude, we have derived a new lower bound for the capacity of the nonlinear fiber channel by taking into account the fact that phase noise is one of the the most significant consequences of nonlinear interference, and by taking advantage of the fact that this noise is characterized by strong temporal correlations. We showed that the peak capacity per polarization can be increased by approximately 0.7bit/s/Hz or equivalently the length of a system can be (almost) doubled for a given transmission rate.

The authors would like to acknowledge financial support from the Israel Science Foundation (grant 737/12). Ronen Dar would like to acknowledge the support of the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and the Yitzhak and Chaya Weinstein Research Institute for Signal Processing. Fruitful discussions with R. J. Essiambre are also gratefully acknowledged.

References

1. P. P. Mitra and J. B. Stark, Nature **411**, 1027–1030 (2001).
2. K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, Physical review letters **91**, 203901–203904 (2003).
3. R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, J. Lightwave Technol. **28**, 662–701 (2010).
4. A. D. Ellis, J. Zhao, and D. Cotter, J. Lightwave Technol. **28**, 423–433 (2010).
5. G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, Opt. Express **19**, B440–B451 (2011).
6. A. Mecozzi and R.-J. Essiambre, J. Lightwave Technol. **30**, 2011–2024 (2012).
7. E. Agrell and M. Karlsson, Optical Fiber Communication Conference 2013 (OFC13).
8. P. J. Winzer and G. J. Foschini, Opt. Express **19**, 16680–96 (2011).
9. R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, Opt. Express **21**, 25685–25699 (2013).
10. T. M. Cover and J. A. Thomas, "*Elements of information theory*,"(Wiley, 1991).
11. R. Dar, M. Shtaif, and M. Feder, European Conference on Optical Communication 2013 (ECOC13), paper P.4.16 (2013).
12. With system parameters similar to those used for evaluating the capacity in [3] and assuming for example a 500km link, the phase has been shown to be nearly constant on the scale of a few tens of symbol durations [9].
13. A. Lapidoth and S. M. Moser, IEEE Trans. on IT **49**, 25685–25699 (2003).