Analytical optimization of wideband nonlinear optical fiber communication systems

ZHENG LIU,[1] TIANHUA XU,[1,2,3,5] 🆔 CENQIN JIN,[2] TONGYANG XU,[3] 🆔 MINGMING TAN,[4] 🆔 JIAN ZHAO,[1,6] AND TIEGEN LIU[1]

[1]*School of Precision Instruments and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China*

[2]*School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom*

[3]*Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, United Kingdom*

[4]*Aston Institute of Photonic Technologies, Aston University, Birmingham B4 7ET, United Kingdom*

[5]*tianhua.xu@ieee.org*

[6]*enzhaojian@tju.edu.cn*

Abstract: In the design of fiber links for both continental and transoceanic optical communication systems, the optimization of span length is of high importance from both performance and cost perspectives. In this work, the maximization of signal-to-noise ratio (SNR) is investigated by optimizing the span length in wideband (up to 4.5-THz) Nyquist-spaced optical fiber communication systems. A simple and accurate closed-form expression of the optimal span length is provided, and a quick estimation of SNR is also described for practically feasible and cost-effective span length values.

1. Introduction

Currently, over 95% of digital data traffic is carried over optical fiber networks [1]. Advanced techniques such as wavelength division multiplexing (WDM), polarization multiplexing, high-order quadrature amplitude modulation (QAM) and multi-channel digital back-propagation (MC-DBP) have been employed to meet the drastically increased demand on the data volume [2–8]. Such systems have high efficiencies of frequency spectrum, while they either need complex hardware signal processing or are sensitive to system parameters and impairments [9,10]. It is thus very important to study how to improve the performance, in order to design optical-fiber communication systems rationally.

Ultra long-haul terrestrial and submarine communication systems, require high and stable performance in terms of signal transmission. With these considerations, designers need to leave enough margin to cope with the random and burst interferences. Therefore, it is of high importance to know where is the optimal case and how much the corresponding performance can achieve. The gap between the current system performance and its optimal value is significantly useful. Designers need to comprehensively analyze these data to design appropriate systems, taking into account both the system performance and the economic cost.

Many research works have been carried out on evaluating the performance of high capacity optical fiber communication systems. The accuracy of the analytical model of the nonlinear interference i.e., enhanced Gaussian noise (EGN) model, has been demonstrated in many reported works [11–27]. The optimization of the information rate, the span length and the signal power has also been numerically and experimentally investigated [27,28]. However, there were no simple and closed-form optimization results provided. This leaves some space for improvement. For example, it is intractable to obtain an analytical form of the optimal span length via the full EGN

model due to its complex structure. The GN model and some other NLI models are simple but not accurate enough [11,16,20,22]. The approximated EGN model [12–15,21] is accurate enough without being over complicated, while the numerical solution using the this model function images requires a lot of computation resources. This is still inefficient for designers. Therefore, these reported works could not be easily applied to the design of optical systems with different transmission parameters.

In this paper, we presented an analytical formula to calculate the optimal span length based on the approximated EGN model and mathematical approximations. We examine the signal-to-noise ratio (SNR) as an indicator of the performance when the transmission span length is varied in practical WDM optical communication systems considering the transceiver noise. In the process of deriving the formula, a simple closed-form estimation of the system SNR is provided. The accuracy of those formulas is also investigated. With those formulas, designers can estimate system the maximum SNR quickly with the transmission parameters, to determine whether the design of the system is feasible. Finally, The SNR is compared for digital signal process (DSP) schemes with electronic dispersion compensation (EDC) at optimized span length and MC-DBP at commonly used span length (80 km) [29] in dual-polarization quadrature phase shift keying (DP-QPSK), dual-polarization 16-ary quadrature amplitude modulation (DP-16QAM), DP-64QAM and DP-256QAM systems to indicate the improvement via the optimization.

This paper is arranged as follows. The numerical transmission setup is described in Section 2. The EGN model is explained in Section 3. Section 4 details the analyses of theoretical formulas and analytical results. Conclusions are drawn in Section 5.

2. Transmission system

Figure 1 illustrates the transmission setup of the dual-polarization multi-channel Nyquist-spaced optical system using modulation formats of DP-QPSK, DP-16QAM, DP-64QAM, and DP-256QAM. In the transmitter, the laser comb lines are modulated with the in-phase and quadrature (I-Q) modulators.
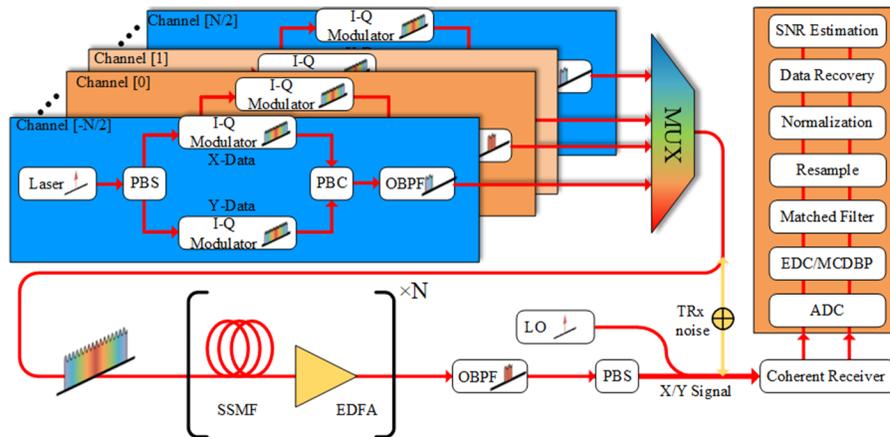
Fig. 1. Schematic of the dual-polarization multi-channel Nyquist-spaced optical fiber communication system. PBS: polarization beam splitter, PBC: polarization beam combiner, LO: local oscillator, ADC: analogue-to-digital convertor, OBPF: optical band-pass filter.

The transmitted symbol sequences in each channel and polarization are fully independent and random. The Nyquist pulse shaping (NPS) is implemented by a root-raised cosine (RRC) filter with a roll-off of 0.1%. The standard single mode fiber (SSMF) is simulated based on the split-step Fourier solution of Manakov equation with a logarithmic step-size distribution [30].

The erbium-doped optical fiber amplifiers (EDFA) are periodically placed to compensate for the loss in the optical fiber. To realize a coherent detection of all in-phase and quadrature signal components in each polarization, the received signals are mixed with an ideal free-running local oscillator (LO) laser carrier at the receiver end. In DSP modules, a frequency domain equalizer is used to implement the EDC [31,32]. The MC-DBP is realized using the reverse operation of the split-step Fourier solution of the Manakov equation with the nonlinearity compensation implemented in the middle of each segment [6,33]. For MC-DBP, an ideal RRC filter is used to select the desired back-propagated bandwidth. Meanwhile, the filter can also remove the unwanted out-of-band amplified spontaneous emission (ASE) noise. After that, a matched filter is used to select the center channel and to remove cross talk from neighboring channels. In total, over 2^{16} symbols are employed to calculate the SNR. The phase noise, the frequency offset of the transmitter and LO lasers, and the differential group delay (DGD) between the two polarizations in the fiber are all neglected [34,35]. Table 1 details parameters used in the numerical simulations.

Table 1. Transmission System Parameters

Parameter	Value
Center wavelength (both transmitter and LO)	1550 nm
Roll-off	0.1%
Attenuation coefficient (α)	0.2 dB/km
Chromatic dispersion coefficient (D)	17 ps/nm/km
Nonlinear coefficient (γ)	1.2 /W/km
EDFA noise figure	6 dB
TRx noise figure	25 dB

3. EGN model

Impairments in optical communication system can be divided into linear distortions and nonlinear distortions [11–13,26,27]. For dual-polarization multi-span EDFA amplified Nyquist-spaced WDM transmission systems without any phase noise. Linear distortions contain ASE noise and transceiver noise (TRx noise). Nonlinear distortions include signal-signal (S-S) interaction and signal-noise (S-N) interaction. Therefore, signal-to-noise ratio can be expressed as follows [16,36]:

$$SNR = \frac{P}{\sigma^2_{TRx} + \sigma^2_{totASE} + \sigma^2_{s-s} + \sigma^2_{s-n}} \tag{1}$$

where ASE noise can be considered as additive Gaussian noise [37], and TRx noise power is proportional to signal power P.

$$\sigma^2_{TRx} = \tilde{\kappa}P \tag{2}$$

$$\sigma^2_{totASE} = N_s\sigma^2_{ASE} = N_s(G-1)F_nh\nu_0R_s \tag{3}$$

where $\tilde{\kappa}$ is TRx noise figure, N_s is the amount of fiber spans, G is EDFA gain, F_n is EDFA noise figure, h is Plank constant, ν_0 is the center lightwave frequency and R_s is the symble rate. From [38–40], the nonlinear noise can be evaluated by the following formulas [14–16]:

$$\sigma^2_{s-s} = N_s^{\epsilon+1}\eta P^3 \tag{4}$$

$$\sigma^2_{s-n}(ASE) = 3\xi_{ASE}\eta\sigma^2_{ASE}P^2 \tag{5}$$

$$\sigma^2_{s-n}(TRx) = 3\xi_{TRx}\eta\tilde{\kappa}P^3 \tag{6}$$

$$\epsilon = \frac{3}{10} log \left[1 + \frac{6}{L_s} \frac{L_{\text{eff}}}{\sinh^{-1}(\frac{\pi^2}{2}|\beta_2|R_s^2 N_{\text{ch}}^2 L_{\text{eff}})} \right] \tag{7}$$

where $L_{\text{eff}} = \frac{1-e^{-\alpha L_s}}{\alpha}$, α is fiber attenuation coefficient, β_2 is second-order dispersion coefficient, N_{ch} is the amount of WDM channels, L_s is the fiber span length, ϵ is the coherence factor, η and ξ are the nonlinear interference (NLI) coefficient and the distance-dependence factor derived by the EGN model [12,41] respectively.

$$\xi_{\text{ASE}} = \sum_{n=1}^{N_s} n^{\epsilon+1} \approx \frac{N_s^{\epsilon+1}}{2} + \frac{N_s^{\epsilon+2}}{\epsilon + 2} \tag{8}$$

$$\xi_{\text{TRx}} = N_s^{\epsilon+1} \tag{9}$$

$$\eta \approx \frac{8}{27} \frac{\alpha \gamma^2 L_{\text{eff}}}{\pi |\beta_2| R_s^2} \sinh^{-1} \left(\frac{\pi^2}{2} |\beta_2| L_{\text{eff}} N_{\text{ch}}^2 R_s^2 \right) - \frac{80}{81} \frac{\kappa \gamma^2 L_{\text{eff}}^2}{\pi |\beta_2| L_s R_s^2} \left[\Phi \left(\frac{N_{\text{ch}} + 1}{2} \right) + C + 1 \right] \tag{10}$$

where the constant κ is related to the fourth standardised moment (kurtosis) of the input signal constellation. Values for different modulation formats are shown in Table 2[26,36]. Function $\Phi(x)$ is the digamma function, $C \approx 0.557$ is the Euler-Mascheroni constant and γ is the fiber nonlinear coefficient.

As for the system with MC-DBP, the S-S interaction will be compensated, depending on the amount of MC-DBP channels. Meanwhile, σ_{s-s}^2 can be expressed as following [17–25]:

$$\sigma_{s-s}^2 = N_s^{\epsilon+1}[\eta(N_{\text{WDM}}) - \eta(N_{\text{MC-DBP}})]P^3 \tag{11}$$

The accuracy of the EGN model has been well tested in previous works [16,18,23,36]. The EGN model is a perfect tool for estimating system SNR. However, sometimes engineers prefer to know the optimal span length or the difference between the actual case and the optimal case at given system parameters. Therefore, a closed-form optimal span length formula is derived based on the EGN model in this work.

Table 2. The value of κ

Modulation Format	Value
DP-QPSK	1
DP-16QAM	$\frac{17}{25}$
DP-64QAM	$\frac{13}{21}$
DP-256QAM	$\frac{121}{200}$

4. Results and discussions

4.1. Simpler SNR estimation

Analyzing the influence of different terms in the EGN model on the signal SNR and considering some mathematical approximations, the following simplified form can be obtained:

$$SNR = -10 \log_{10} \left[\tilde{C} \frac{e^{\alpha L_s}}{\alpha L_s} + \tilde{\kappa}_1 (\alpha L)^{Ax} x^{1+A} - \tilde{\kappa}_2 x^2 + \tilde{\kappa} \right] \tag{12}$$

$$\tilde{C} = \frac{h \nu_0 R_s F_n L \alpha}{P} \tag{13}$$

$$\tilde{\kappa}_1 = \frac{8}{27} \frac{P^2}{\alpha} C_1 C_0 (\alpha L)^{\left(1 + \frac{4.2}{C_1}\right)\left(1 + \frac{6}{C_1}\right)^{-0.7}} \tag{14}$$

$$\tilde{\kappa}_2 = \frac{80}{81}\kappa L P^2 C_0 \Phi(\frac{N_{ch}+1}{2}) \tag{15}$$

$$C_0 = \frac{\gamma^2}{\pi|\beta_2|R_s^2} \tag{16}$$

$$C_1 = \log(\frac{\pi^2|\beta_2|R_s^2 N_{ch}^2}{\alpha}) \tag{17}$$

$$A = \frac{3}{10}\frac{6}{C_1}(1 + \frac{6}{C_1})^{-\frac{7}{10}} \tag{18}$$

where $x = \frac{L_{\text{eff}}}{L_s}$, L is the communication distance and $\tilde{\kappa}$ is TRx noise figure. For the detailed derivation process, please refer to Appendix A. There are some numbers in Eq. (14) and Eq. (18) that arise from fitting the ϵ (Eq. (7)) with a biased first-order Taylor expansion, and they are constant. Moreover, A is only related to the square of bandwidth, and the change is small when the bandwidth is over 2 THz. Therefore, it can be further simplified by using a constant to replace related parameters. The accuracy of Eq. (12) has been examined and the result is shown in Fig. 2.
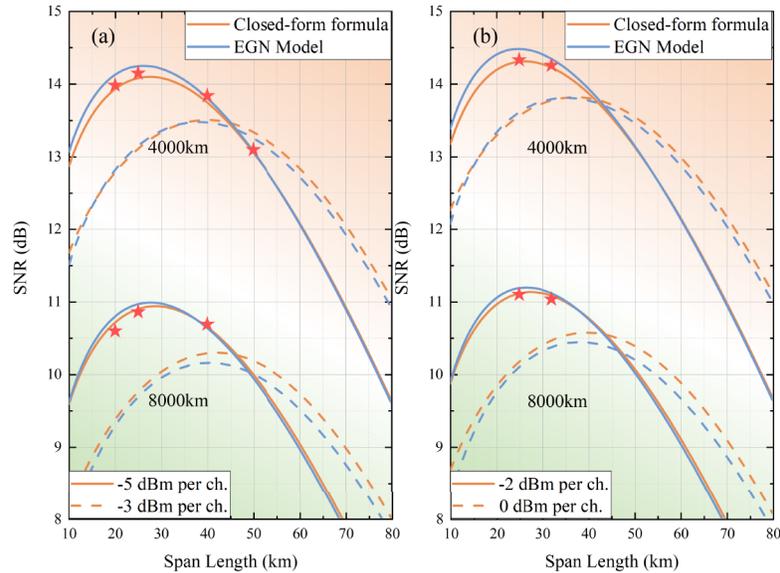
Fig. 2. The accuracy of closed-form formula in Eq. (12), compared to the EGN model. (a) 61 channels and 32 GBaud per channel. (b) 33 channels and 64 GBaud per channel. Red dots are simulation results.

Figure 2 shows the difference in terms of predicted SNR between Eq. (12) and the EGN model. The transmission distances are 4000 km in the orange area and 8000 km in the green area, respectively. The red dots in Fig. 2 are simulation results. It is shown that the performance of system working at the optimal power (-5 dBm in Fig. 2(a) and -2 dBm in Fig. 2(b), solid line) and the performance of system operating at a higher signal power (-3 dBm in Fig. 2(a) and 0 dBm in Fig. 2(b), dashed line). It can be seen that the discrepancies between Eq. (12) and the EGN model are always smaller than 0.3 dB, in terms of SNRs. These deviations are within the acceptable range in practical applications. Besides, the results from the approximated EGN model are also very close to the simulation results.

4.2. Optimal span length estimation

Using the simplified EGN model (Eq. (12)) and corresponding mathematical approximation, the formula for the optimal span can be obtained as follows:

$$L_s(\text{Opt.}) = \log\left(\frac{5.8\tilde{C} + \tilde{\kappa}_1(0.13\log(\alpha L)+1) - 2\tilde{\kappa}_2}{1.3\tilde{C}}\right)/0.076$$

$$\tilde{\kappa}_1 = \frac{8}{27}\frac{P^2}{\alpha}C_0 C_1(\alpha L)^{\left(1+\frac{6}{C_1}\right)^{0.3}} \quad \tilde{\kappa}_2 = \frac{80}{81}\kappa L P^2 C_0 \Phi\left(\frac{N_{ch}+1}{2}\right) \tag{19}$$

$$C_0 = \frac{\gamma^2}{\pi|\beta_2|R_s^2} \quad C_1 = \log\left(\frac{\pi^2|\beta_2|R_s^2 N_{ch}^2}{\alpha}\right) \quad \tilde{C} = \frac{h\nu_0 R_s F_n L\alpha}{P}$$

Appendix B shows the derivation procedure in detail.

Figure 3 shows the difference of the estimated optimal span length from Eq. (19) and from the EGN model at the bandwidth of half C-band (~ 2 THz), including four groups of curves (a)-(d). The EDFA noise figure is 6 dB in all four groups. Fig. 4 shows the difference of the estimated optimal span length from Eq. (19) and from the EGN model at the bandwidth of C-band (~ 4.5 THz), include three groups of curves (a)-(c). Group (c) has a lower EDFA noise figure (4.5 dB) than the group (a) and the group (b) (6 dB). All systems work near the optimal signal power. It can be seen that the discrepancy decreases when the transmission length goes larger and the difference is always within 5%.
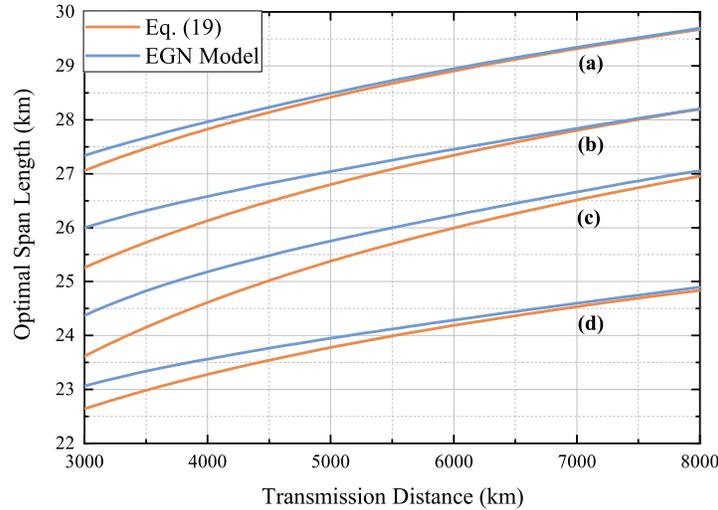
Fig. 3. The optimal span length v.s. transmission distance. (a) 128 GBaud 15 channels at 1 dBm per channel; (b) 64 GBaud 33 channels at -2 dBm per channel; (c) 32 GBaud 61 channels at -5 dBm per channel; (d) 48 GBaud 51 channels at -4 dBm per channel. EDFA noise figure is 6 dB in all curves.

Besides, it can be found that when the EDFA noise figure (F_n) grows up, the optimal span length $L_s(\text{Opt.})$ goes down. That indicates that the choice of EDFAs with lower noise figure could decrease the system costs, since the optimal span length increases. Although the unit price of EDFAs may increase, fewer EDFAs will be used due to the extension of the optimal span length. This would save total costs in long-haul communication systems. Taking an 8000 km communication system as an example, the use of EDFAs with a noise figure of 4.5 dB rather than with a noise figure of 6 dB could reduce the total costs if the unit price of EDFA is increased by less than 12%. On the other hand, the SNR performance will improve due to the decrease of EDFA noise figure. This means that the system has the potential to increase the data rate (e.g. by using higher-order modulation formats) in the future.
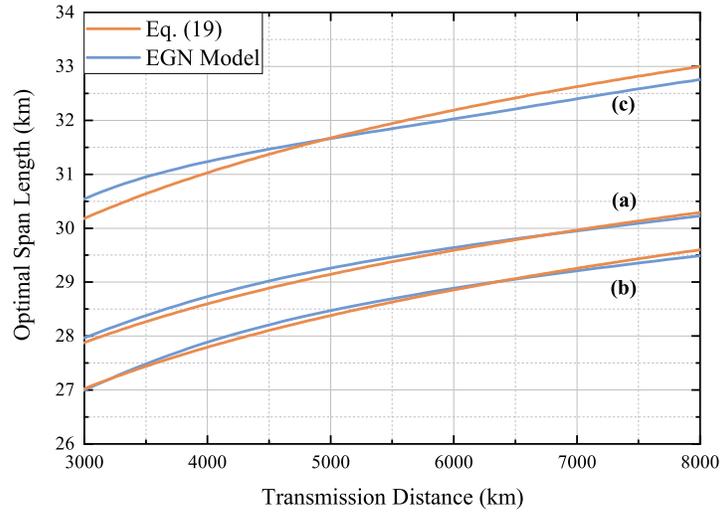
Fig. 4. The optimal span length versus the transmission distance. (a) 64 GBaud 71 channels at -2 dBm per channel, EDFA noise figure is 6 dB; (b) 32 GBaud 141 channels at -5 dBm per channel, EDFA noise figure is 6 dB; (c) 64 GBaud 71 channels at -2 dBm per channel, EDFA noise figure is 4.5 dB

4.3. Analysis of optimal span length considering commercial factors

The optimal span length described above corresponds to the best sysmtem performance (SNR) with specific transmission schemes. It represents the fiber span length when the maximum SNR can be achieved using specific EDFA, modulation format, and symbol rate, etc. Usually, the calculated optimum fiber span length is relatively short ($\leqslant$ 35 km). In this section, we use an indicator of cost/bit to analyze the optimal span length taking into account commercial factors, and the corresponding cost parameters are shown in the Table 3 [42]. The difference in C_{OA} of two optical amplifiers with different noise figures (4.5 dB and 6 dB) comes from Ref. [43,44].

Table 3. System cost parameters (Normalized)

Modulation Format	Value
Deployment per km (C_D)	0.7
Cable per km (C_C)	0.5
Fiber per km (C_F)	0.005
Optical amplifier (6 dB) (C_{OA})	2
Optical amplifier (4.5 dB) (C_{OA})	2.3
Transponder per 100G (C_T)	1

The system cost is calculated by $C_{tot} = (C_D + C_C + C_F M)L + C_{OA}NM + C_T C$, where C_D, C_C, C_F, C_{OA} and C_T are listed in Table 3, M is the number of spatial paths and its value is 16 (8 fiber pairs) in this paper, L is transmission distance, N is the amount of EDFAs and C represents the capacity which calculated by $C = 2MB \log_2(1 + SNR)$ and B is the system bandwidth (4.5 THz) [42].

Figure 5 shows the relationship between the cost/bit (or the capacity) and the fiber span length. The solid lines correspond to the capacity and the dash lines to the cost/bit. It can be found that the calculated optimal span length to maximize the capacity is smaller than the estimated optimal

span length to achieve the minimum (best) cost/bit value. Therefore, considering the commercial factors only, the optimal span length is near 60 km. This will lose some system capacity and will cause the communication system to work sub-optimally. The use the optimal span length derived by Eq. (19) can allow the maximum capacity, but it will cost more to build such systems. Meanwhile, considering the blue lines and the orange lines in Fig. 5(a) or Fig. 5(b), it can be found that the use of EDFAs with a lower noise figure not only can improve the system capacity, but also can decrease the cost per bit value. Taking Fig. 5(b) as an example, if the capacity requirement is 475 Tb/s in the 6 dB EDFA scheme, the fiber span length is near 38 km. The cost/bit is 0.048 correspondingly. For the 4.5 dB EDFA scheme, in order to achieve the same or higher capacity requirement, the fiber span length can be selected as between 13 km and 60 km. It can be found that when the span length is between 40 km and 60 km, the 4.5 dB EDFA transmission scheme not only improves the system capacity but also reduces the cost/bit. Besides, as mentioned in Section 4.2, it may also decrease the system total costs by replacing EDFAs with a lower noise figure because of the longer fiber spans. Therefore, the use of EDFAs with lower noise figures is a solution to balance the performance and the costs.
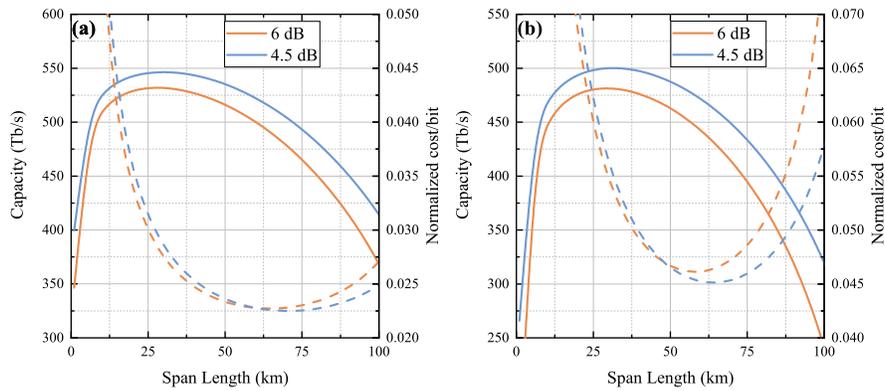
Fig. 5. Capacity and cost per bit versus fiber span length at the transmission distances of 6000 km (a) and 11000 km (b). EDFA noise figure is 6 dB in orange lines and 4.5 dB in blue lines.

4.4. Fast estimation of commonly used span length

The system SNR can be calculated in a simpler way, than the use of Eq. (12), when span length is higher than 50 km.

Some mathematical approximations are detailed in Appendix C. The final formula is shown as follows:

$$SNR = -10 \log 10 \left[\tilde{C} \exp(\alpha L_s)(\alpha L_s)^{-1} + \kappa_1 (\alpha L_s)^{-1} - \kappa_2 (\alpha L_s)^{-2} + 0.05\kappa_1 + \tilde{\kappa} \right]$$
$$\tilde{C} = \frac{h\nu_0 R_s F_n L \alpha}{P}$$
$$\kappa_1 = \frac{8}{27} \frac{P^2}{\alpha} \frac{\gamma^2}{\pi |\beta_2| R_s^2} \log\left(\frac{\pi^2 |\beta_2| R_s^2 N_{ch}^2}{\alpha} \right)$$
$$\kappa_2 = \frac{80}{81} \kappa L P^2 \frac{\gamma^2}{\pi |\beta_2| R_s^2} \Phi\left(\frac{N_{ch}+1}{2} \right)$$

$$(20)$$

Fig. 6 shows the difference between Eq. (20) and the EGN model for different span lengths, and the accuracy is very high.
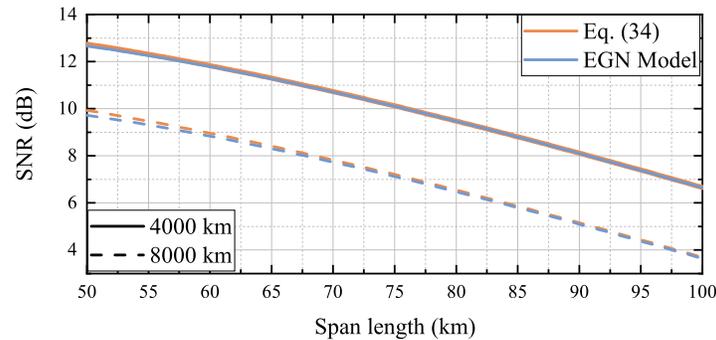
Fig. 6. The accuracy of Eq. (20) over 50 km per span at the bandwidth of 2 THz.

4.5. *Improvement from the use of the optimal span length*

Through Eq. (19), the optimal SNR of the system can be easily and quickly calculated, as shown in Fig. 7 orange lines. It has been known that the SNRs of systems with 16QAM, 64QAM and 256QAM behave close to each other [27], so we only plotted the 16QAM case in Fig. 7(a) and Fig. 7(c). To demonstrate the improvement from the use of the optimal span length, Fig. 7 draws the SNR of the system with 80 km per span (purple lines). It can be seen that compared to the 80 km/span system, the improvement is significant. For the submarine optical communication system, the span length has been set to 50 km to cope with the harsh communication environment (dark lines) [29]. However, 50 km is still larger than the optimal span length. The SNR can be further improved by about 1 dB via the reduction of fiber span length as shown in light green area in Fig. 7. In addition, we have compared the improvement from the use of MC-DBP and the improvement from the optimization of span length. Currently, the digital coherent receiver has a maximum bandwidth of 300 GHz [45] that means only 9-channel MC-DBP for 32 GBaud system and 5-channel MC-DBP for 64 GBaud can be achieved. Interestingly, the SNR in the system with 300 GHz MC-DBP (80km per span) behaves similar to the system with 50km per span (EDC only). This means that when the bandwidth of the coherent receiver and the speed of the DSP cannot make a breakthrough, the optimization span length has always been an effective option to improve the system performance.

5. **Conclusion**

The EGN model can estimate the system SNR accurately, however, the computational complexity is significant and the optimal solution of span length can not be obtained easily. In this work, we have simplified EGN model with some mathematical approximations and proved its accuracy. After that, the analytical formula of the optimal span length is derived by solving simplified formulas. We have derived a fast and accurate formula to estimate the optical fiber span length to maximize SNR of the optical communication systems. We have already examined its accuracy and it can reduce the computational complexity significantly compared to the EGN model. Meanwhile, we analysed the optimal span length not only in terms of performance but also with respect to commercial factors. To balance the cost and the signal performance, the noise figure of optical amplifiers is also an important factor. Generally, the EDFAs with lower noise figure have more commercial benefits especially in systems with longer transmission distances. For systems with a fiber span length over 50 km, we have further simplified the EGN model and the simplified formula only needs to calculate exponential function once and multiplication three times. Finally, the improvement using optimal span lengths is examined compared to the use of

MC-DBP. Our work provides a simple and analytical solution at the design of the optical fiber links in both continental and transoceanic optical communication systems.
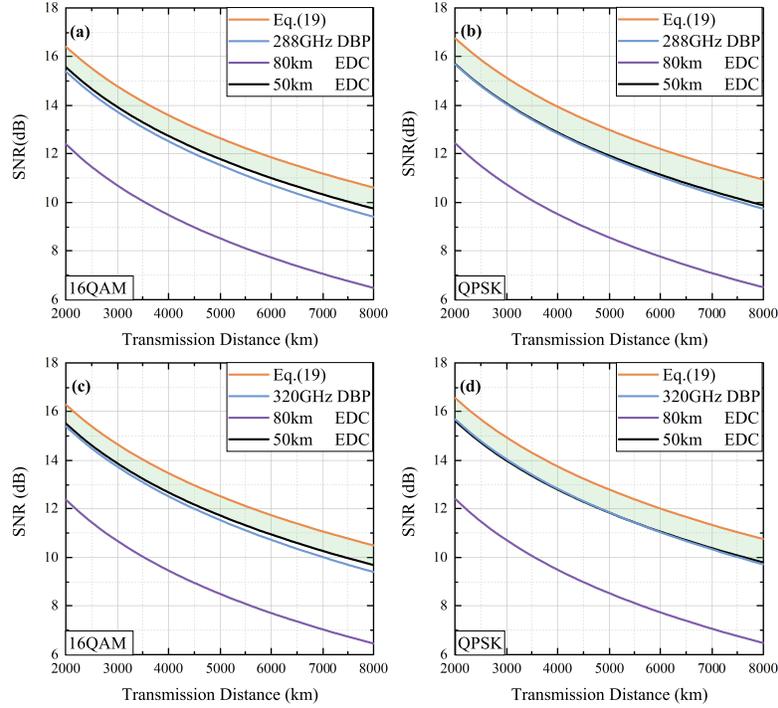
Fig. 7. The SNR at different transmission schemes. Orange lines: the theoretical optimal span length calculated by Eq. (19). Blue lines: 80 km per span using MC-DBP. Purple lines: 80 km per span EDC only. Dark lines: 50 km per span EDC only. (a) and (b) 32 GBaud 61 channels at optimal launched power; (c) and (d) 64 GBaud 33 channels at optimal launched power.

Appendices

Appendix A. SNR estimation

For a communication system without MC-DBP, its SNR can be calculated using Eq. (1) - Eq. (10). To simplify the formula, we first ignore some terms that contribute little to SNR. Consider the nonlinear noise term, it can be found that they have the same pre-factor,

$$\sigma_{s-s}^2 + \sigma_{s-n}^2 = N_s^{\epsilon+1}\eta P^3 \left[1 + 3(\frac{1 + N_s}{2}\frac{\sigma_{ASE}^2}{P} + \tilde{\kappa})\right] \tag{21}$$

where $\xi_{ASE} \approx \frac{N_s^{\epsilon+1} + N_s^{\epsilon+2}}{2}$, $\tilde{\kappa}$ is TRx noise figure. When N_s is big enough, the second term in brackets in Eq. (21) approximately equals $1.5\frac{\sigma_{totASE}^2}{P}$, which nearly equals 0 because the signal power P is far greater than ASE noise power σ_{totASE}^2. Besides, $\tilde{\kappa}$ is 25 dB ≈ 0.0032 is negligible. Therefore,

$$SNR \approx \frac{P}{\sigma_{TRx}^2 + \sigma_{totASE}^2 + \sigma_{s-s}^2} \tag{22}$$

From the EGN model and the normalization of the numerator in Eq. (22), we could get

$$SNR = \left[N_s \frac{\sigma_{ASE}^2}{P} + (N_s^{\epsilon+1}\eta_{GN} - N_s\eta_{corr})P^2 + \tilde{\kappa} \right]^{-1} \tag{23}$$

$$\eta_{GN} = \frac{8}{27} \frac{\alpha\gamma^2 L_{eff}}{\pi|\beta_2|R_s^2} \sinh^{-1}\left(\frac{\pi^2}{2}|\beta_2|L_{eff}N_{ch}^2 R_s^2 \right) \tag{24}$$

$$\eta_{corr} = \frac{80}{81} \frac{\kappa\gamma^2 L_{eff}^2}{\pi|\beta_2|L_s R_s^2} \Phi\left(\frac{N_{ch}+1}{2} \right) \tag{25}$$

where η_{GN} represents the η derived by the GN model and η_{corr} is the correction term derived by the EGN model. When x is big enough, $\sinh^{-1}(x) = \log(x + \sqrt{1+x^2}) = \log(2x)$. Further, $\sinh^{-1}\left(\frac{\pi^2}{2}|\beta_2|L_{eff}N_{ch}^2 R_s^2 \right) \approx \log\left(\frac{\pi^2|\beta_2|N_{ch}^2 R_s^2}{\alpha} \right)$ is a constant. Let C represent it, and C_0 represent $\frac{\gamma^2}{\pi|\beta_2|R_s^2}$. Substitute $L = N_s L_s$ into Eq. (23) and use $\frac{L_{eff}}{L_s}$ as an independent variable x.

$$SNR = \left[N_s \frac{\sigma_{ASE}^2}{P} + LP^2(N_s^\epsilon \kappa_1 x - \kappa_2 x^2) + \tilde{\kappa} \right]^{-1} \tag{26}$$

where $\kappa_1 = \frac{8}{27}C_0 C$, $\kappa_2 = \frac{80}{81}C_0\kappa\Phi(\frac{N_{ch}+1}{2})$, and L is the transmission distance. From the EGN model [16], ϵ approaches 0, and optimal span is usually small meaning that x approaches 1 [28]. N_s is finite and ϵ changes little, so it is feasible to fit corresponding curves with linear functions. As follows,

$$SNR = -10\log_{10}\left[\tilde{C}\frac{e^{\alpha L_s}}{\alpha L_s} + \tilde{\kappa_1}(\alpha L)^{Ax}x^{1+A} - \tilde{\kappa_2}x^2 + \tilde{\kappa} \right] \tag{27}$$

$$\tilde{C} = \frac{h\nu_0 R_s F_n L\alpha}{P} \tag{28}$$

$$\tilde{\kappa_1} = \frac{8}{27}\frac{P^2}{\alpha}C_1 C_0(\alpha L)^{\left(1+\frac{4.2}{C_1}\right)\left(1+\frac{6}{C_1}\right)^{-0.7}} \tag{29}$$

$$\tilde{\kappa_2} = \frac{80}{81}\kappa LP^2 C_0\Phi(\frac{N_{ch}+1}{2}) \tag{30}$$

$$C_0 = \frac{\gamma^2}{\pi|\beta_2|R_s^2} \tag{31}$$

$$C_1 = \log(\frac{\pi^2|\beta_2|R_s^2 N_{ch}^2}{\alpha}) \tag{32}$$

$$A = \frac{3}{10}\frac{6}{C_1}(1 + \frac{6}{C_1})^{-\frac{7}{10}} \tag{33}$$

where $\tilde{\kappa}$ is TRx noise figure.

Appendix B. Optimal span length estimation

The problem of solving the optimal span can be transformed into an optimization problem. Based on Eq. (12) - Eq. (18), we need to solve the following formula to calculate the optimal span length

$$\arg\min \tilde{C}\frac{e^{\alpha L_s}}{\alpha L_s} + \tilde{\kappa}_1(\alpha L)^{Ax}x^{1+A} - \tilde{\kappa}_2 LP^2 x^2 \tag{34}$$

It can be proved that Eq. (34) has a single solution, therefore we only need to set the derivative of Eq. (34) to 0 to obtain the optimal L_s. Define $T = A\log(\alpha L)$ and the function can be expressed as follows,

$$\tilde{C}\frac{1}{x}\frac{d}{dx}\left(\frac{e^{\alpha L_s}}{\alpha L_s}\right) + \tilde{\kappa}_1 e^{Tx}x^{A-1}(Tx + 1 + A) - 2\tilde{\kappa}_2 = 0 \tag{35}$$

Eq. (35) can be divided into two parts. The first part is only related to the fiber loss coefficient (α) and the fiber span length (L_s). We use an exponential function with weight and bias to fit $\frac{1}{x}\frac{d}{dx}\left(\frac{e^{\alpha L_s}}{\alpha L_s}\right)$

$$\frac{1}{x}\frac{d}{dx}\left(\frac{e^{\alpha L_s}}{\alpha L_s}\right) = \frac{1}{x}\frac{d}{dx}\left(\frac{1}{x}e^{\frac{1}{x}} + g(x)\right) = \frac{1}{x}(1 - \frac{1}{x})e^{\frac{1}{x}} + \frac{1}{x}g'(x) \tag{36}$$

the optimal span length is usually small within 30 km, and x is close to 1. The fit curve of function $\frac{1}{x}(1 - \frac{1}{x})e^{\frac{1}{x}}$ can be fitted as follows,

$$\tilde{C}\frac{1}{x}\frac{d}{dx}\left(\frac{e^{\alpha L_s}}{\alpha L_s}\right) = \tilde{C}(5.8 - 1.3e^{0.076L_s}) \tag{37}$$

Eq. (37) is accurate enough when L_s within 20-50 km as long as the fiber loss coefficient is 0.2 dB/km. As for other type of fibers, the fitting parameters are shown in Table 4.

Table 4. Fitting parameters

Fiber types	Bias	Weight	Scale
SSMF ($\alpha = 0.2dB/km$)	5.8	-1.3	0.076
ULL ($\alpha = 0.16dB/km$)	7.3	-1.2	0.067

Now we consider the second part of Eq. (35). In the process of deriving Eq. (12), due to some approximations, the optimal point has been shifted to the right hand side. It is noted that the value of the second part of Eq. (35) changes very slowly when L_s is close to the optimal value. To derive the analytical formula, we use a value near its minimum as an approximate representation which can also compensate for the error caused by the shift of the optimal point. Finally, the closed-form optimal span length formula can be expressed as follows:

$$L_s(\text{Opt.}) = \log\left(\frac{5.8\tilde{C} + \tilde{\kappa}_1(0.13\log(\alpha L) + 1) - 2\tilde{\kappa}_2}{1.3\tilde{C}}\right)/0.076$$

$$\tilde{\kappa}_1 = \frac{8}{27}\frac{P^2}{\alpha}C_0 C_1(\alpha L)^{\left(1 + \frac{6}{C_1}\right)^{0.3}} \quad \tilde{\kappa}_2 = \frac{80}{81}\kappa LP^2 C_0\Phi(\frac{N_{ch} + 1}{2}) \tag{38}$$

$$C_0 = \frac{\gamma^2}{\pi|\beta_2|R_s^2} \quad C_1 = \log(\frac{\pi^2|\beta_2|R_s^2 N_{ch}^2}{\alpha}) \quad \tilde{C} = \frac{h\nu_0 R_s F_n L\alpha}{P}$$

Appendix C. Fast estimation formula

For the terms include x in Eq. (12)

$$x = \frac{L_{\text{eff}}}{L_s} = \frac{1 - e^{-\alpha L_s}}{\alpha L_s} \approx (\alpha L_s)^{-1} \tag{39}$$

Therefore,

$$Ax = \frac{A}{\alpha L_s} \approx 0, \quad \tilde{\kappa}_1 (\alpha L)^{Ax} x^{1+A} = \tilde{\kappa}_1 (\alpha Ls)^{-(1+A)} \tag{40}$$

The $(\alpha L_s)^{-(1+A)}$ is still complicated, and A is a number that changes very little with bandwidth

$$d(\alpha L_s)^{-(1+A)} = (\alpha L_s)^{-1} \log(\alpha L_s) dA \approx 0.05 \tag{41}$$

when $dA \approx 0.13$ and $(\alpha L_s)^{-1} \log(\alpha L_s) \approx 0.36$. Therefore, $(\alpha L_s)^{-(1+A)} \approx (\alpha L_s)^{-1} + 0.05$. Besides, note that Eq. (14) can also be further simplified

$$\left(1 + \frac{4.2}{C_1}\right)\left(1 + \frac{6}{C_1}\right)^{-0.7} \approx 0 \tag{42}$$

Therefore,

$$SNR = -10 \log 10 \left[\tilde{C} \exp(\alpha L_s)(\alpha L_s)^{-1} + \kappa_1 (\alpha L_s)^{-1} - \kappa_2 (\alpha L_s)^{-2} + 0.05\kappa_1 + \tilde{\kappa}\right]$$

$$\tilde{C} = \frac{h\nu_0 R_s F_n L\alpha}{P}$$

$$\kappa_1 = \frac{8}{27} \frac{P^2}{\alpha} \frac{\gamma^2}{\pi |\beta_2| R_s^2} \log\left(\frac{\pi^2 |\beta_2| R_s^2 N_{ch}^2}{\alpha}\right) \tag{43}$$

$$\kappa_2 = \frac{80}{81} \kappa L P^2 \frac{\gamma^2}{\pi |\beta_2| R_s^2} \Phi\left(\frac{N_{ch}+1}{2}\right)$$

References

1. "Cisco visual networking index (vni) complete forecast update, 2017-2022, https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1213-business-services-ckn.pdf. 2018.
2. B. Mukherjee, *Optical WDM networks* (Springer Science & Business Media, 2006).
3. D. Qian, N. Cvijetic, J. Hu, and T. Wang, "108 Gb/s OFDMA-PON with polarization multiplexing and direct detection," J. Lightwave Technol. **28**(4), 484–493 (2010).
4. T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for *M*-QAM constellations," J. Lightwave Technol. **27**(8), 989–999 (2009).
5. J. Lu, K. B. Letaief, J.-I. Chuang, and M. L. Liou, "M-PSK and M-QAM BER computation using signal-space concepts," IEEE Trans. Commun. **47**(2), 181–184 (1999).
6. G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, "On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission," Opt. Express **22**(24), 30053–30062 (2014).
7. R. Dar and P. J. Winzer, "On the limits of digital back-propagation in fully loaded WDM systems," IEEE Photonics Technol. Lett. **28**(11), 1253–1256 (2016).
8. C. Fougstedt, L. Svensson, M. Mazur, M. Karlsson, and P. Larsson-Edefors, "Asic implementation of time-domain digital back propagation for coherent receivers," IEEE Photonics Technol. Lett. **30**(13), 1179–1182 (2018).
9. A. Ellis, M. McCarthy, M. Al Khateeb, M. Sorokina, and N. Doran, "Performance limits in optical communications due to fiber nonlinearity," Adv. Opt. Photonics **9**(3), 429–503 (2017).
10. T. Xu, N. A. Shevchenko, Y. Zhang, C. Jin, J. Zhao, and T. Liu, "Information rates in Kerr nonlinearity limited optical fiber communication systems," Opt. Express **29**(11), 17428–17439 (2021).

11. P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," J. Lightwave Technol. **32**(4), 694–721 (2013).
12. P. Poggiolini, "A generalized GN-model closed-form formula," arXiv:1810.06545 [eess.SP] (2018).
13. M. Ranjbar Zefreh, A. Carena, F. Forghieri, S. Piciaccia, and P. Poggiolini, "A GN/EGN-Model Real-Time Closed-Form Formula Tested over 7, 000 Virtual Links, arXiv e-prints arXiv:1911.03328 (2019).
14. M. R. Zefreh, F. Forghieri, S. Piciaccia, and P. Poggiolini, "Accurate closed-form model for nonlinear fiber propagation supporting both high and near-zero dispersion regimes," Opt. Express **29**(7), 10825–10852 (2021).
15. M. R. Zefreh and P. Poggiolini, "A GN-model closed-form formula supporting ultra-low fiber loss and short fiber spans," arXiv:2111.04584 [eess.SP] (2021).
16. P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of nonlinear propagation in uncompensated optical transmission links," IEEE Photonics Technol. Lett. **23**(11), 742–744 (2011).
17. A. Carena, G. Bosco, V. Curri, P. Poggiolini, M. T. Taiba, and F. Forghieri, "Statistical characterization of pm-qpsk signals after propagation in uncompensated fiber links," in *36th European Conference and Exhibition on Optical Communication*, (IEEE, 2010), pp. 1–3.
18. M. R. Zefreh, F. Forghieri, S. Piciaccia, and P. Poggiolini, "A closed-form nonlinearity model for Forward-Raman-Amplified WDM optical links," in *Optical Fiber Communication Conference (OFC) 2021*, (Optical Society of America, 2021), p. M5C.1.
19. M. R. Zefreh and P. Poggiolini, "Characterization of the link function in GN and EGN methods for nonlinearity assessment of ultrawideband coherent fiber optic communication systems with raman effect," arXiv:2009.12687 [eess.SP] (2020).
20. P. Serena and A. Bononi, "On the accuracy of the Gaussian nonlinear model for dispersion-unmanaged coherent links," in *European Conference & Exhibition on Optical Communication*, (2013).
21. P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "A simple and effective closed-form GN model correction formula accounting for signal Non-Gaussian distribution," J. Lightwave Technol. **33**(2), 459–473 (2015).
22. P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," J. Lightwave Technol. **30**(24), 3857–3879 (2012).
23. Z. Liu, T. Xu, J. Ding, Y. Zhang, M. Li, T. Xu, and T. Liu, "Accuracy of EGN model in ultra-wideband optical fiber communication systems," in *Photonics Asia*, vol. 11891 (SPIE, 2021), pp. 31–37.
24. N. A. Shevchenko, T. Xu, D. Semrau, G. Saavedra, and P. Bayvel, "Achievable information rates estimation for 100-nm Raman-Amplified optical transmission system," in *Ecoc European Conference on Optical Communication*, (2016).
25. P. Serena and A. Bononi, "A time-domain extended Gaussian noise model," J. Lightwave Technol. **33**(7), 1459–1472 (2015).
26. N. A. Shevchenko, T. Xu, D. Lavery, L. Gabriele, D. J. Ives, R. I. Killey, and B. Polina, "Modeling of nonlinearity-compensated optical communication systems considering second-order signal-noise interactions," Opt. Lett. **42**(17), 3351 (2017).
27. T. Xu, N. A. Shevchenko, D. Lavery, S. Daniel, L. Gabriele, A. Alex, R. I. Killey, and B. Polina, "Modulation format dependence of digital nonlinearity compensation performance in optical fibre communication systems," Opt. Express **25**(4), 3311 (2017).
28. B. Karanov, T. Xu, N. A. Shevchenko, D. Lavery, R. I. Killey, and P. Bayvel, "Span length and information rate optimisation in optical transmission systems using single-channel digital backpropagation," Opt. Express **25**(21), 25353–25362 (2017).
29. K. Zhong, X. Zhou, Y. Wang, T. Gui, Y. Yang, J. Yuan, L. Wang, W. Chen, H. Zhang, J. Man, L. Zeng, C. Yu, A. P. T. Lau, and C. Lu, "Recent advances in short reach systems," in *Optical Fiber Communication Conference*, (Optical Society of America, 2017), p. Tu2D.7.
30. G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "Suppression of spurious tones induced by the split-step method in fiber systems simulation," IEEE Photonics Technol. Lett. **12**(5), 489–491 (2002).
31. R. Kudo, T. Kobayashi, K. Ishihara, Y. Takatori, A. Sano, and Y. Miyamoto, "Coherent optical single carrier transmission using overlap frequency domain equalization for long-haul optical systems," J. Lightwave Technol. **27**(16), 3721–3728 (2009).
32. T. Xu, G. Jacobsen, S. Popov, J. Li, and Y. Zhang, "Chromatic dispersion compensation in coherent transmission system using digital filters," Opt. Express **18**(15), 16243 (2010).
33. R. Asif, C.-Y. Lin, M. Holtmannspoetter, and B. Schmauss, "Optimized digital backward propagation for phase modulated signals in mixed-optical fiber transmission link," Opt. Express **18**(22), 22796–22807 (2010).
34. A. Bononi, R. Dar, M. Secondini, P. Serena, and P. Poggiolini, "Fiber nonlinearity and optical system performance," in *Springer Handbook of Optical Networks* (Springer, 2020), pp. 287–351.
35. C. Jin, N. A. Shevchenko, Z. Li, S. Popov, Y. Chen, and T. Xu, "Nonlinear coherent optical systems in the presence of equalization enhanced phase noise," J. Lightwave Technol. **39**(14), 4646–4653 (2021).
36. A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," Opt. Express **22**(13), 16335–16362 (2014).
37. P. G. Agrawal, *Fiber-optic communication systems* (Wiley, 2004).

38. T. Tanimura, M. Nölle, J. K. Fischer, and C. Schubert, "Analytical results on back propagation nonlinear compensator with coherent detection," Opt. Express **20**(27), 28779–28785 (2012).

39. D. Rafique and A. D. 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 (2011).

40. L. Beygi, N. V. Irukulapati, E. Agrell, P. Johannisson, M. Karlsson, H. Wymeersch, P. Serena, and A. Bononi, "On nonlinearly-induced noise in single-channel optical links with digital backpropagation," Opt. Express **21**(22), 26376–26386 (2013).

41. M. R. Zefreh, F. Forghieri, S. Piciaccia, and P. Poggiolini, "A closed-form nonlinearity model for Forward-Raman-Amplified WDM optical links," in *Optical Fiber Communication Conference (OFC) 2021*, (Optical Society of America, 2021), p. M5C.1.

42. R. Dar, P. J. Winzer, A. R. Chraplyvy, S. Zsigmond, K.-Y. Huang, H. Fevrier, and S. Grubb, "Cost-optimized submarine cables using massive spatial parallelism," J. Lightwave Technol. **36**(18), 3855–3865 (2018).

43. "Multicom 1550 EDFA key features," https://www.multicominc.com/wp-content/uploads/MUL-EDFA-V-1-X-EDFA-Flyer_Spec-OCT-2017.pdf (2017).

44. "FS 20dBm Output In-line DWDM EDFA amplifier, 25dB gain, LC/UPC," https://img-en.fs.com/file/datasheet/m6200-20la-datasheet.pdf (2021).

45. K. Shi, E. Sillekens, and B. C. Thomsen, "246 GHz digitally stitched coherent receiver," in *Optical Fiber Communication Conference*, (Optical Society of America, 2017), p. M3D.3.