

# The error statistics analysis of the QPSK-modulated signal in the high-rate optical link

Alexey Redyuk, Anton Skidin, Michael Fedoruk

*Institute of Computational Technologies, Novosibirsk, 6 Acad. Lavrentiev avenue,  
630090, Russia*

*Novosibirsk State University, Novosibirsk, 2 Pirogova street, 630090, Russia*

Alex Shafarenko

*University of Hertfordshire, Department of Computer Science, Hatfield, AL10 9AB, UK*

---

## Abstract

A model of an optical fibre communication line operating at 80 Gb/s with QPSK modulation format is studied numerically. The error statistics and patterning effects are analyzed and nonlinear transmission impairments of the optical signal are quantified.

*Keywords:*

optics communications, direct detection, phase modulation, modelling, nonlinear effects

---

## 1. Introduction

The steadily growing traffic transmitted through the modern optical communication lines stimulates studies of new methods to increase the fiber line carrying capacity. There are several ways to achieve this in wave-division-multiplexing (WDM) systems. One way is to increase the number of frequency channels. Alternatively one could increase the data rate of the channels. The increase of a channel data rate can be implemented using either

---

*Email addresses:* [alexey.redyuk@gmail.com](mailto:alexey.redyuk@gmail.com) (Alexey Redyuk), [ask@skidin.org](mailto:ask@skidin.org) (Anton Skidin), [mife@ict.nsc.ru](mailto:mife@ict.nsc.ru) (Michael Fedoruk), [A.Shafarenko@herts.ac.uk](mailto:A.Shafarenko@herts.ac.uk) (Alex Shafarenko)

shorter carrier pulses or modulation formats that have more degrees of freedom (bits) per carrier [1]. Such high level modulation formats are especially useful when the signal is being processed using the coherent detection approach ([2]-[3]). When the number of channels increases, the total power is typically also increased; the higher power exacerbates the nonlinear effects in the fiber. In particular, they manifest themselves as uneven error statistics (i.e. patterning effects), as they make the bit-error rate (*BER*) dependent on the data being transmitted through the fiber ([4]-[10]). The statistical knowledge of error distributions can be used to improve the error correction schemes ([6]-[10]). Patterning effects can occur due to a variety of causes, e.g. intrachannel fourwave mixing for OOK-modulated signal ([11]-[15]).

Therefore, there is a clear interest in exploration of patterning effects and the nonlinear impairments to the signal as a function of the signal power for various modulation formats. Although the modern QAM modulation formats involve both amplitude and phase keying, amplitude patterning was encountered already in the simplest OOK formats. By contrast the patterning effects that affect the phase are less studied and less understood, therefore we focus here on those in the sequel.

## 2. System Configuration

Fig. 1 defines the transmission link used in our analysis. The unit cell of the link contains a two-step dispersion map with a span of single-mode fibre (SMF) 85 km long followed by a span of dispersion-compensated fibre (DCF). The accumulated dispersion of the SMF is fully compensated by the DCF, so the residual dispersion of a single unit cell is set to 0 ps/nm. The fibre parameters and simulation parameters are shown in Table 1 and Table 2 respectively. The gain of either erbium-doped fibre amplifier (EDFA) is 13.2 dB, which ensures that the losses are fully compensated in each unit cell. The noise figure of the amplifiers is set to 4.5 dB.

At the transmitter, each input symbol is converted into a pulse of a certain waveform. The waveform and power for all launch pulses are the same, whereas the phase varies. Fig. 2 shows the constellation diagram for quadrature phase-shift keying with Gray coding, whereby any two adjacent symbols only differ by one bit. For simulation we used the following raised cosine waveform of a launch pulse:  $f(t) = 0.5(1 + \cos(2\pi t/T))$ , where  $T = 25$  ps is the duration of the symbol interval,  $t \in [-T/2; T/2]$ . In this paper we limit ourselves to input pulses with a launch peak power of  $P_0 = 10$  mW.

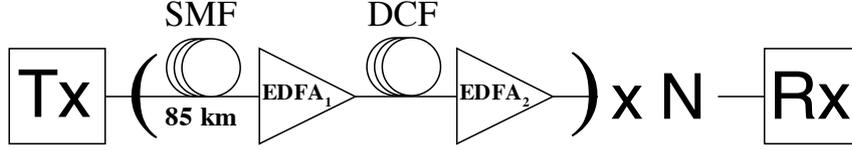


Figure 1: The scheme of the transmission link.

Table 1: Fibre parameters

	SMF	DCF
Attenuation at 1550 nm	0.20 dB/km	0.65 dB/km
Dispersion	17 ps/nm/km	-100 ps/nm/km
Dispersion slope	0.07 ps/nm <sup>2</sup> /km	-0.41 ps/nm <sup>2</sup> /km
Nonlinear parameter	1.37 W <sup>-1</sup> km <sup>-1</sup>	5.76 W <sup>-1</sup> km <sup>-1</sup>
Length	85 km	14.45 km

Table 2: Simulation parameters

Symbol rate	40 Gbaud
Bit rate	80 Gbit/s
Sequence length	2 <sup>15</sup>
Samples/symbol	32
Carrier wave frequency	1550 nm
Distance	100 km x N spans; N=1..30

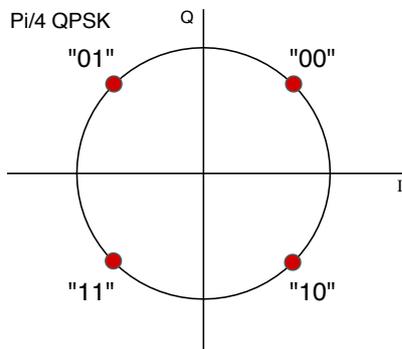


Figure 2: Constellation diagram for QPSK with gray coding.

Table 3: Relation between phase and symbol

$0 < \varphi \leq \pi/2$	"00"
$\pi/2 < \varphi \leq \pi$	"01"
$\pi < \varphi \leq 3\pi/2$	"11"
$3\pi/2 < \varphi \leq 2\pi$	"10"

Here the propagation of a train of pulses with those parameters across the link is calculated using the Split Step Fourier Transform (SSFT) technique [16, 17].

At the receiver end a Gaussian optical band pass filter with  $B_{FWHM} = 110$  GHz is used. The procedure of detection is as follows: we assume that using coherent detection one can obtain the real and imaginary parts of the output signal  $A(t)$ . There are 32 samples for each output pulse and we calculate the phase  $\varphi = \text{arctg}(\text{Im}A(t)/\text{Re}A(t))$  in the middle point of the symbol interval. Noise and nonlinear effects cause distortion of the pulse leading to some distortion of its phase. Next using the knowledge of the phase in the middle point of the symbol interval, we find what symbol corresponds to the received pulse. The correspondence between phase and symbol is shown in Table 3. After the output symbol sequence has been worked out it is compared with the input one.

Nonlinear effects are known to cause an additional phase drift. The drift accumulates as the signal propagates along the fiber. Fig. 3 shows the

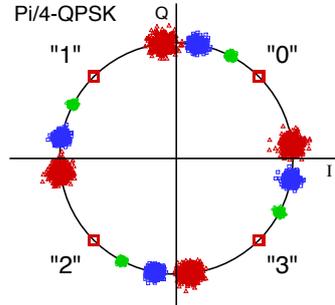


Figure 3: Constellation diagram of phase drift for QPSK. Red square - initial state, green clouds - after 1 unit cell, blue clouds - after 2 unit cells, red clouds - after 3 unit cells.

constellation diagram for a QPSK-signal, which exemplifies the phase drift on a sequence of random symbols propagating along the fiber.

The magnitude of the phase drift is directly proportional to the signal power and the transmission distance. Consequently the phase drift may be a major cause of errors at the receiver. Accordingly, most practical BPSK fibre links apply differential phase keying. However, if we use QPSK with coherent detection we need to perform phase recovery. This can be done at the receiver end using digital signal processing (DSP).

### 3. Results

We have done 50 numerical experiments to study the propagation of a QPSK-signal along the transmission link described above. A pseudorandom sequence of  $2^{15}$  symbols was used in each simulation. Direct detection of the signal was performed at the end of each unit cell using the model of receiver described above and the results have been averaged over all 50 experiments.

The symbol sequences at the output were partitioned into (overlapping) triplets of consecutive symbols. The triplets were then classified into correct and erroneous based on how the middle symbol was decoded. QPSK uses 4 different points on the constellation diagram, so there are  $4 \times 4 \times 4 = 64$  different triplets. However all triplets can be divided into classes within which they are indistinguishable in terms of phase relationships. For example, triplets "000", "111", "222" and "333" have the same phase profile shifted

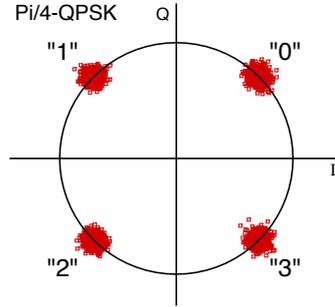


Figure 4: Constellation diagram after compensation of phase drift for QPSK.

Table 4: Total numbers of errors for 16 basic classes

000&111&222&333	364	16.79%	020&131&202&313	352	16.24%
001&112&223&330	96	4.43%	021&132&203&310	121	5.58%
002&113&220&331	13	0.60%	022&133&200&311	16	0.74%
003&110&221&332	98	4.52%	023&130&201&312	57	2.63%
010&121&232&303	17	0.78%	030&101&212&323	17	0.78%
011&122&233&300	92	4.24%	031&102&213&320	64	2.95%
012&123&230&301	350	16.14%	032&103&210&321	294	13.56%
013&120&231&302	124	5.72%	033&100&211&322	93	4.29%

by  $\pi/2$  relative to one another. As a result, one can reduce the whole variety to only 16 different classes. Table 4 shows the number of erroneous triplets detected for each of the 16 classes after the distance of 1400 km. The total number of errors is equal to 2168. This corresponds to a *SER* of  $10^{-3}$ .

From table 4 it is obvious that the error statistics is uneven. Some triplets greatly contribute to the total number of errors and some triplets almost do not produce erroneous symbols.

Triplets “000” as the worst one, “002” as the best one and “001” as medium one were examined in detail to explain the uneven distribution of error statistics. A pseudorandom sequence of  $2^{15}$  symbols was transmitted over the simulated line. The amplifiers were made noiseless in order to allow the effect of fibre nonlinearity alone to be assessed. Fig. 5 shows the eye diagrams for the triplets “000”, “001” and “002”.

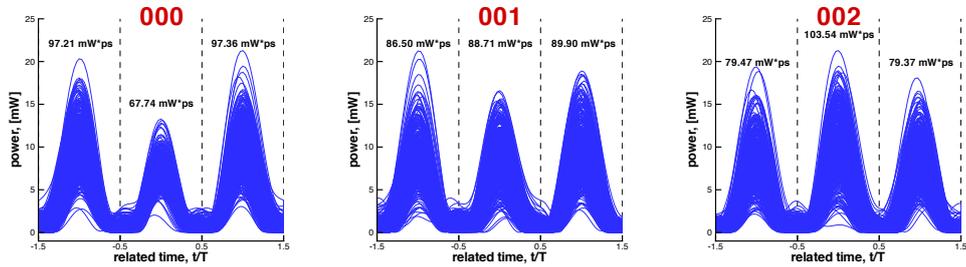


Figure 5: Optical intensity eye diagrams for triplets “000”, “001” and “002”.

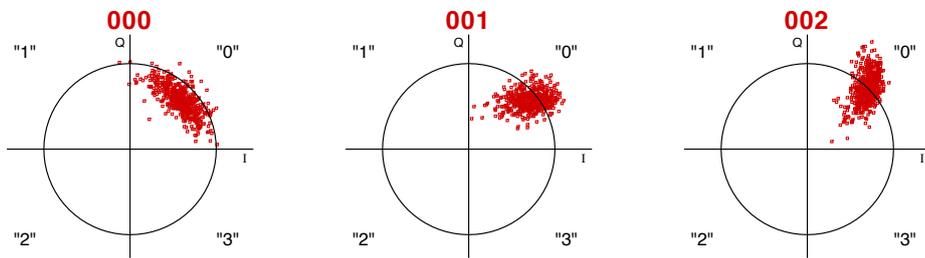


Figure 6: Constellation diagrams for triplets “000”, “001” and “002”.

The figure appears to suggest that the power of the middle pulse of the triplet “000” decreases whereas the power of the end pulses increases. The average energy of the pulses was calculated to confirm this. For the triplet “002” we observed a power rise in the middle pulse and a similar drop in the end pulses.

Fig. 6 shows constellation diagrams for the triplets “000”, “001” and “002”. Each point of a cloud in this diagram corresponds to the real and imaginary parts of the complex amplitude of the middle pulse calculated in the middle of the symbol interval. At this point the receiver decodes the symbol.

As follows from figure 6, the constellation cloud of the triplet “000” is stretched along the circumference whereas the constellation cloud of the triplet “002” is stretched along the radius. This explains the anomalously high error count for the triplet “000”: a constellation cloud stretching along the circle begins to move outside the catchment area of the symbol earlier

than a constellation cloud stretching along the radius. It is worth noting that most points for the triplet “000” are inside the circle. It indirectly confirms that for the triplet “000” the average energy of the middle pulse is less than that of its neighbor pulses. The constellation cloud of the triplet “002” has a slight twist and, the greater the pulse amplitude, the stronger the twist. The triplet “001” is neither the worst nor the best according to Table 4 and its constellation cloud has a third state, which is in between stretching along the circle and the radius.

#### 4. Error statistics ~~behaviour~~

As shown above, the error statistics for 1400 km ~~was~~ uneven. In general, the type of error statistics ~~can allow (or cannot allow)~~ to reduce the bit-error rate using the special encoding schemes developed for that purpose ([10, 9]). ~~So, it is useful to estimate the degree of “unevenness” of the statistics for the different transmission distances, and for different signal powers.~~

First provide the formal definition of the degree of “unevenness” for the error statistics. As ~~the~~ <sup>a</sup> measure of that, we ~~will~~ <sup>propose to</sup> use the per-bit entropy. Suppose, we have the triplet error statistics (similar to one shown in table 4). If we know that the receiver symbol is transmitted with an error, we can consider the error statistics as the information source that can give us the information about the error origin. So, the entropy of that information source represents the uncertainty in the error source. Particularly, if we have an uneven error statistics, its per-bit entropy will be 1 (maximum uncertainty).

In the case of QPSK triplets we can estimate the error statistics per-bit entropy using the following equation:

$$E = \frac{-\sum_{i=1}^{4^3} P_i \cdot \log_2 P_i}{\log_2 4^3} = -\frac{1}{6} \cdot \sum_{i=1}^{64} P_i \log_2 P_i,$$

where  $P_i$  is the probability that the error in the message was caused by the triplet of number  $i$ . The multiplier  $1/6$  normalizes the entropy to the range from 0 to 1 as if it would be the binary information source.

We have investigated the dependence between OSNR and the error statistics for the initial signal power of 10 mW. The figure 7 shows the obtained results. From the figure we can see that the error statistics is approaching to the uniform statistics (the entropy of uniform error distribution is 1) when the OSNR decreases and the distance increases. So, the special encoding

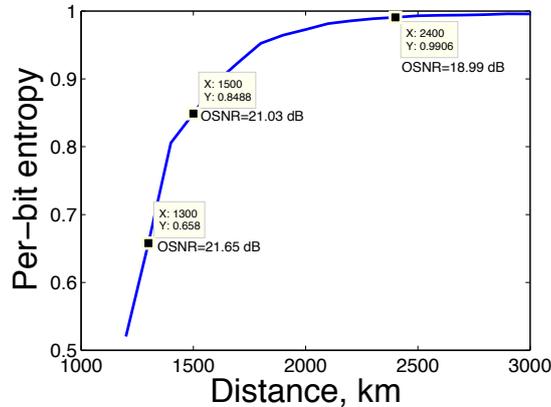


Figure 7: The dependence between the error statistics entropy and the distance.

methods can be applied only if OSNR is sufficient large to provide an uneven error distribution. The per-bit entropy represents the uncertainty of the error source. When this value is 1, the error source is absolutely unpredictable (this is a case of uniform error statistics); in contrast, if the entropy is 0, the error sources is absolutely predictable (there is either no error at all or all errors have the same origin, for instance they observed in the same triplet).

We have investigated the behaviour of the error statistics for the initial signal power of 1 mW. The OSNR has varied from 22.8 dB (after 100 km) to 8 dB (after 3000 km). This fact confirms the hypothesis that when the OSNR is small, the error statistics tends to have the uniform distribution.

It should be noted that even if the error statistics is strongly uneven, it does not guarantee that the special encoding methods will operate effectively, since the code redundancy of those methods depends, beside others, on the symbols the triplet consists of (it is shown, for example, in [10]). However, the evenness of the error statistics guarantees that the employing of the special encoding methods is ineffective.

## 5. Advanced detection scheme

The results described above show the presence of the uneven error statistics for QPSK data triplets. So, there is the opportunity to use special coding schemes to reduce the bit-error rate (those schemes are proposed in [10, 9]).

Table 5: The distinction between average points of the triplet and ideal QPSK points.

Point	Average value for the triplet ( $A$ )	Ideal QPSK point
$P_{000}(1400)$	(1.81, 1.80)	(2.21, 2.21)
$P_{001}(1400)$	(2.28, 1.79)	(2.21, 2.21)
$P_{002}(1400)$	(1.78, 2.31)	(2.21, 2.21)
$P_{020}(2800)$	(1.93, -1.63)	(2.21, -2.21)
$P_{021}(2800)$	(1.90, -1.15)	(2.21, -2.21)
$P_{022}(2800)$	(1.26, -1.78)	(2.21, -2.21)
$P_{023}(2800)$	(1.34, -1.19)	(2.21, -2.21)

However, any coding scheme decreases the effective data rate due to code redundancy. Below we propose irredundant approach to improve the standard middle point detector described above.

The “advanced” detector takes into account the distinction between constellation diagrams for different triplets (those distinctions can be observed on figure 6). Denote as  $x_i$  the symbol that is being processed by our detector. Its neighbours are  $x_{i-1}$  and  $x_{i+1}$ . Also denote as  $Q$  the data point that received at position  $i$ . The goal of the detector is to retrieve  $x_i$  from the phase point  $Q$ .

To obtain  $x_i$ , we can use the position of  $Q$  on the phase plain, compare it with the ideal QPSK points (see figure 2) and decide what  $x_i$  is. However, we can compare the received point position with an another point. The detector proposed below uses the comparison with the average points of the triplets  $x_{i-1}0x_{i+1}$ ,  $x_{i-1}1x_{i+1}$ ,  $x_{i-1}2x_{i+1}$ ,  $x_{i-1}3x_{i+1}$  and  $x_i$  has the value  $s$ , if the average point of the triplet  $x_{i-1}sx_{i+1}$  is nearest to the point  $Q$ .

To adjust the detector to the optimal performance, we need to find the average points for each data triplet. They were found for any distance using the results of numerical experiments. So, we have for distance  $d$  64 points on the phase plain ( $P_1(d)$ ,  $P_2(d)$ , ...,  $P_{64}(d)$ ). To highlight the distinction between them and the ideal QPSK points, some points are given in the table 5.

The detection scheme works as follows:

1. The symbols  $x_{i-1}$  and  $x_{i+1}$  are detected using the standard middle-point detector described above.

2. The detector selects four points:  $S_0 = P_{x_{i-1}0x_{i+1}}$ ,  $S_1 = P_{x_{i-1}1x_{i+1}}$ ,  $S_2 = P_{x_{i-1}2x_{i+1}}$ ,  $S_3 = P_{x_{i-1}3x_{i+1}}$ .

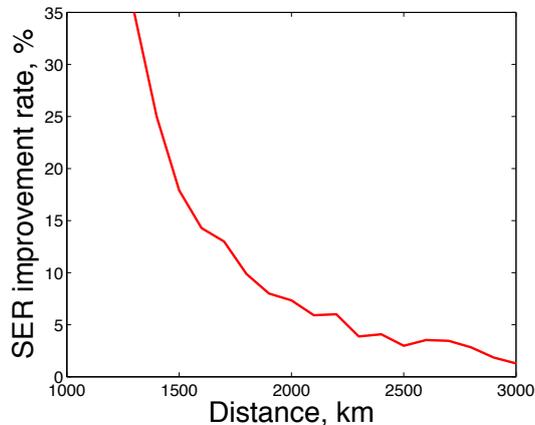


Figure 8: The BER reduction rate obtained using the advanced detection scheme.

3. The detector estimates the values:  $\rho_0 = |Q - S_0|$ ,  $\rho_1 = |Q - S_1|$ ,  $\rho_2 = |Q - S_2|$ ,  $\rho_3 = |Q - S_3|$ .  $\rho_j$  is the distance between  $Q$  and  $S_j$ .

4.  $x_i = \{j | \rho_j = \min\{\rho_0, \rho_1, \rho_2, \rho_3\}\}$ .

This scheme was examined for different transmission distances. On the figure 8 the results are presented. We can see that using proposed detection scheme, the error reduction of about 10–30% was obtained for the distances of 1000-1700 km. When the distance reaches the limit of 2000 km, the error statistics becomes uniform (see figure 7) and the advanced detection does not have any considerable improvement.

## 6. Discussion

Modern optical fibre communication lines use primarily QAM formats [18, 19, 20] as QAM has high spectral efficiency. However, even though the properties of QAM transmission are well understood in linear transmission regimes, the patterning effects in the nonlinear affected transmission change the behaviour quite significantly. We have already mentioned in the introduction that amplitude patterning had been studied previously [21, 9]. There are some newer results regarding differential BPSK [7] where phase behaviour was essential for correct decoding. However, nonbinary representations, and especially those that involve both the amplitude and the phase, present a challenge to statistical analysis as the number of possible patterns increase

exponentially with the number of levels of phase and amplitude. It appears essential to study low grade PSK formats first to learn qualitative features of patterning as it occurs in simple nonbinary cases with a constant amplitude, of which the first one to consider is QPSK.

In our study we discovered that despite the simplicity of QPSK, in non-linear mode interactions between phase and amplitude are already manifest. Furthermore, there is a clear qualitative pattern of behaviour, whereby certain phase relationships in a triplet promote amplitude scattering whilst others primarily affect the phase and yet a third category appears to exhibit no significant constellation anisotropy of any kind. We do not yet have a theoretical model to explain why some phase relationships lead to one pattern and some to another. However, already the empirical knowledge gleaned from our observations is useful as it can inform signal encoding/decoding.

There are two ways in which the encoding/decoding logic could be so informed. Firstly, our results may influence constrained pre-encoding [14, 6, 10]. The constrained code removes the symbol patterns that are prone to decoding errors at the receiver at the expense of code redundancy. Experiments with constrained coding in a patterned binary channel have already been carried out (see [9]), and the results demonstrate a significant reduction of the *BER* (up to 50% for a 5%-redundant code). In [9] it was shown that the method is applicable for nonlinear channels with a large *BER* ( $BER > 5 \cdot 10^{-2}$ ). In such a case the constrained code can reduce the error rate prior to the forward error correction stage in order to restore the efficiency of the FEC, which is essential when the frequency of errors approaches critical levels at which the FEC is likely to fail.

The other way in which the knowledge of error statistics can be used is with automatic classification methods [22]. The statistics would be useful for tuning the classifier to a particular pattern of signal behaviour. Classifiers continually adjust a partitioning of the parameter space in which code symbols are identified with individual partitions, which is known as training. The quality of a classifier depends on how many parameters have to be fitted; any a priori knowledge of the behavioural statistics speeds up the training and improves classification.

## 7. Conclusions

We have studied the effects of nonlinearity of optical fibre transmission on the signal phase when it is used in QPSK format. In this research we

disregarded amplitude keying in order to study the cross-interaction between phase and amplitude in symbol triplets in isolation. We discovered that constellation clouds respond with certain types of anisotropy to specific phasing structures of the triplet. **We proposed an advanced detection scheme that allows to reduce the bit-error rate using the irredundant algorithm that does not affect the effective data rate and provide the BER improvement from 10 to 30 percents at middle transmission distances. Also we discussed other methods used to improve the signal transmission quality that based on the channel coding and classification schemes.** Future research will focus on how stable those patterns are with respect to amplitude coding in triplets and on specific pre-encoding methods that are likely to be effective in mitigating these classes of nonlinear impairments.

## 8. Acknowledgements

The work was supported by grant No 11.519.11.4018 of the Ministry of Education and Science of the Russian Federation. The authors are grateful to S.P. Hunt for his assistance with the experimental side of their work.

## References

- [1] Winzer P.J., Essiambre R.-J. Proceedings of the IEEE 94, 952, 2006.
- [2] Okoshi T., Kikuchi K. Coherent Optical Fiber Communications, Tokyo: KTK, 1988.
- [3] Derr F. Electronics Letters 27(23), 2177, 1991.
- [4] Settembre M., Matera F., Hagele V., et al. J. of Lightwave Technology 15 (6), 962, 1997.
- [5] Shapiro E. G., Fedoruk M. P., Turitsyn S. K. Electronics Letters 37 (19), 1179, 2001.
- [6] Shafarenko A., Turitsyn K. S., Turitsyn S. K. IEEE Trans. Comm. 55(2), 237, 2007.
- [7] Turitsyn S.K., Fedoruk M.P., Shtyrina O.V. et al. Optics Comm. 277(2), 264, 2007.

- [8] Slater B., Boscolo S., Shafarenko A., Turitsyn S. K. *J. Opt. Networking* 6 (8), 984, 2007.
- [9] Skidin A., Redyuk A., Shtyrina O. et al. *Optics Comm.* 284(19), 4695, 2011.
- [10] Shafarenko A., Skidin A., Turitsyn S. K. *IEEE Trans. Comm.*, 58(10), 2845, 2010.
- [11] Essiambre R.-J., Mikkelsen B., Raybon G. *Electronics Letters* 35(18), 1576, 1999.
- [12] Mamyshev P.V., Mamysheva N.A. *Optics Letters* 24(21), 1445, 1999.
- [13] Shake I., Takara H., Mori K. et al. *Electronics Letters* 34(16), 1600, 1998.
- [14] Rao V. S., Djordjevic I. B., Vasic B. *IEE Proc. Optoelectronics*, 153(2), 87, 2005.
- [15] Shapiro E.G., Fedoruk M.P., Turitsyn S.K., Shafarenko A. *IEEE PTL* 15(10), 1473, 2003.
- [16] Hasegawa A., Tappert F. *Appl. Phys. Letters* 23, 142, 1973.
- [17] Agrawal G.P. *Fiber-Optic Communication Systems* (New York: John Wiley & Sons, Inc., 1997).
- [18] Qian D. et al. 101.7-Tb/s (370×294-Gb/s) PDM-128QAM-OFDM Transmission over 3 × 55-km SSMF using Pilot-based Phase Noise Mitigation, *OFC/NFOEC*, 2011.
- [19] Zhou X. et al. 64-Tb/s, 8 b/s/Hz, PDM-36QAM Transmission Over 320 km Using Both Pre- and Post-Transmission Digital Signal Processing // *Journal of Lightwave Technology*, Vol. 29(4), Pp. 571–577, 2011.
- [20] Sano A. et al. 69.1-Tb/s (432 × 171-Gb/s) C- and Extended L-Band Transmission over 240 Km Using PDM-16-QAM Modulation and Digital Coherent Detection // *OFC 2010, PDPB7*.
- [21] Shapiro E.G., Fedoruk M.P., Turitsyn S.K. *Electronics Letters*, 40(22), 1436, 2004.

- [22] Hunt S.P. et al. Adaptive Electrical Signal Postprocessing with Varying Representations in Optical Communication Systems, Proceedings of the 11th International Conference on Engineering Applications of Neural Networks (EANN2009), pp. 235-245, 2009.