Efficient Weakly-Constrained Codes for Mitigation of Patterning Effects in Digital Communications

Alexander Shafarenko[*], Michael Fedoruk[†], Anton Skidin[‡] and Sergei K. Turitsyn[§]

[*]Department of Computer Science, University of Hertfordshire,
College Lane, Hatfield, AL10 9AB, UK.
E-mail: comqas@herts.ac.uk

[†]Institute of Computational Technologies SB RAS,
Novosibirsk, 6 Acad. Lavrentjev avenue, 630090, Russia.
Novosibirsk State University, Novosibirsk, 2 Pirogova Str., 630090, Russia.
E-mail: mife@ict.nsc.ru

[‡]Institute of Computational Technologies SB RAS,
Novosibirsk, 6 Acad. Lavrentjev avenue, 630090, Russia.
E-mail: askidin@gmail.com

[§]Aston University, Photonics Research Group, Birmingham B4 7ET, UK.
E-mail: s.k.turitsyn@aston.ac.uk

Abstract—We propose specific modulation block codes for weakly-constrained coding and demonstrate that their performance is close to the theoretical limit. We also show that when such codes are used for the mitigation of patterning effects in optical fibre communications, a gain of about 1 dB is possible under realistic conditions, if only at the expense of small redundancy ($\approx 10\%$).

Index Terms—patterning effects, weakly-constrained codes, optical communication, information theory, coding theory

I. INTRODUCTION

The patterning effect due to inter-symbol interference (ISI) is the dependence of the transmission result for one bit of data on the values of its neighbours. ISI is one of the most severe limiting factors in system performance at high data rates. A continuing increase of channel rates with the corresponding decrease of the carrier pulse width and large bit-overlapping during transmission makes effects of ISI even more important for future ultra-high-bit-rate fibre communication systems. In fibre optic communication ISI can result from physical mechanisms of widely varying origins. For instance, the pattern dependence of errors can be caused by the gain saturation of a semiconductor optical amplifer (see, e.g., [1]) or by resonance interactions between pulses in bit-overlapping quasi-linear transmission regimes.

A very important practical example of ISI in high-bit-rate transmission (at channel rates higher than 40 Gbit/s) is the effect of intra-channel four-wave-mixing (ICFWM) [2,3] that leads to the generation of "ghost" pulses that cause performance degradation. In the case of transmission affected by ICFWM, the main contribution to the bit error rate comes from the ghost pulses that appear in logical-zero time slots surrounded by symmetric patterns of logical-ones. Channel pre-encoding has already been considered for mitigation of channel impairments (see e.g. [4-11]) in high-bit-rate systems. In particular, suppression of ICFWM effects by using certain modulation codes was proposed in [5]. Note that advanced FEC coding schemes can also take into account the particular channel error statistics freeing up additional power budget. However, close to the operational regimes where the FEC characteristics get steeper, the transition between error-free performance and performance severely degraded by errors when FEC fails becomes very sharp, which is undesirable from the system design view point. Therefore, even using advanced FEC schemes, it might be practically very important to use channel pre-encoding in the BER feedback loop to reduce the impact of transmission impairments. Below we present modulation block codes for weakly-constrained channel encoding and demonstrate that their performance is close to the theoretical limit.

TABLE I
ERROR STATISTICS FOR 5x40 GBIT/S WDM RZ-DBPSK
SMF/DCF LINE

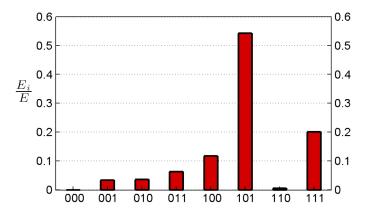| Triplet	Ch. 1	Ch. 2	Ch. 3	Ch. 4	Ch. 5
000	0	2	4	2	6
001	1216	1353	1471	1559	1604
010	7081	7154	7783	8299	8470
011	111	123	181	190	192
100	1064	1250	1286	1422	1464
101	7650	8029	8590	9298	9354
110	94	122	134	145	168
111	42	53	70	60	93

Fig. 1.   Triplet error weights for OOK transmission format

II. EXAMPLES OF PATTERNING EFFECTS IN FIBRE COMMUNICATION SYSTEMS

Recent investigations dedicated to data transmission through optical channels [9,10] indicate that patterning effects take place in optical lines of various structure. For example, in [9] the WDM RZ-DBPSK SMF/DCF channel with hybrid Raman/EDFA amplification was examined. The error statistics for this model are given in Table I.

In [10] a transmission line using the OOK format is described. As shown in [10], the OOK format leads to error statistics in which the 101 triplet error probability is at least 10 times the error probability of any of the other triplets.

In Fig.1 the error statistics for the same line as the one in [9] are given. Here E_i is the error count for the ith triplet and $E = \sum_{i=0}^{7} E_i$ is the total error count. The Fig.1 contains data for the OOK format, and the table statistics are for the DPSK format. Evidently, the 101 triplet makes the largest contribution to BER in any case, which make it important to try to suppress the errors associated with it. This can be done by the skewed encoding scheme, proposed in [7].

III. SKEWED ENCODING AND ITS ASSOCIATED REDUNDANCY

In this section we provide the information theory analysis of the skewed coding. This analysis will be used below.

We define the BER of a received message as follows: $BER = \sum_{k=0}^{7} Q_k R_k$, where Q_k is the average frequency of the triplet k in the input bit string, and R_k is the error probability for the central digit in the triplet (we enumerate triplets according to their binary code, e.g. triplet 6 is 110). The message is assumed to be long enough to neglect the triplets arising on its ends that involve bits outside the message. An uneven distribution of errors R_k caused by the patterning effects makes it possible to reduce the error rate by reducing the frequency of the triplets that affect the BER most. This requires a mapping of the source message onto code words with an uneven (skewed) statistics of triplets, which is a process that we call skewed pre-encoding. Obviously, skewed preencoding can only be achieved at the expense of the information content, which it is convenient to associate with the transmitted signal entropy H (measured in bits/digit). The trade-off here is between the BER improvement, the reduction of the data rate and the increased complexity of the encoder/decoder.

First, we must describe the source information content as a function of the pre-encoding skew. We model the encoder as a general Markov chain shown in Fig.2. The vertices of the graph correspond to the state of the process, which consists of the three most recently transmitted digits including the current one. The transmission of the next digit is depicted as the transition from the current state to the next, keeping the most recent two bits and adjoining to them a new one (either 0 or 1), with a probability that depends on the current state. We use dashed arrows for transitions that produce a zero, and solid ones for those producing a one (see [7] for more detail). All transitions in the diagram generally occur with a probability different from 1/2, which corresponds to a random bit stream without pre-encoding. The difference between the two probabilities from the same state k: $\varepsilon_k = T_{k,2k} - T_{k,2k+1}$, assuming modulo 8 indexing, is a measure of the statistical skew of the encoder.

The stationary distribution corresponding to the Markov process presented in Fig.2 must satisfy the condition: $\hat{T}\vec{Q} = \vec{Q}$, where $\hat{T} = T_{k,l}$ is the process transition matrix graphically described in Fig.2. The solution of this equation normalized by $\sum_{k=0}^{7} Q_k = 1$ gives the frequency of the triplet k in an infinite string

of bits. In the particular case of our channel, the goal is to reduce the frequency of triplet 5 (101) which was found to be the main cause of errors [9]. This can be done by applying $\varepsilon_0 = \varepsilon_1 = \varepsilon_3 = \varepsilon_4 = \varepsilon_5 = \varepsilon_7 = 0$, $\varepsilon_2 = \varepsilon_6 = -\varepsilon$ (cf [7]). Then the probability of the transitions leading to the bad state 101 is $(1 - \varepsilon)/2$, and the probability of the transitions leading to a neighboring "good" state 100 is $(1 + \varepsilon)/2$. In this case the per-bit entropy of the transmitted signal is: $h = \dfrac{3 + g(\varepsilon) + \varepsilon}{4 + \varepsilon}$, and the redundancy $R = \dfrac{1 - g(\varepsilon)}{4 + \varepsilon}$, where $g(\varepsilon) = -\dfrac{1 + \varepsilon}{2} \log_2 \dfrac{1 + \varepsilon}{2} - \dfrac{1 - \varepsilon}{2} \log_2 \dfrac{1 - \varepsilon}{2}$.

Generalizing the result of [5], [7] presents the solution of the general Markov process in a closed form. Since the purpose of encoding is to favor particular states and avoid others, no matter what current state, it may suffice to select the skew parameters in a manner that ignores the source state. In other words we could set $\varepsilon_0 = \varepsilon_4$, $\varepsilon_1 = \varepsilon_5$, $\varepsilon_2 = \varepsilon_6$, and $\varepsilon_3 = \varepsilon_7$. As a result, the general solution of the stationary process reads:

$$
\begin{aligned}
Q_0 &= \frac{(1 - \varepsilon_0)(1 - \varepsilon_2)(1 - \varepsilon_3)}{\Sigma}, \\
Q_1 = Q_4 &= \frac{(1 + \varepsilon_0)(1 - \varepsilon_2)(1 - \varepsilon_3)}{\Sigma}, \\
Q_2 &= \frac{(1 + \varepsilon_0)(1 - \varepsilon_1)(1 - \varepsilon_3)}{\Sigma}, \quad (1) \\
Q_3 = Q_6 &= \frac{(1 + \varepsilon_0)(1 + \varepsilon_1)(1 - \varepsilon_3)}{\Sigma}, \\
Q_5 &= \frac{(1 + \varepsilon_0)(1 + \varepsilon_2)(1 - \varepsilon_3)}{\Sigma}, \\
Q_7 &= \frac{(1 + \varepsilon_0)(1 + \varepsilon_1)(1 + \varepsilon_3)}{\Sigma}.
\end{aligned}
$$

Here the normalizing factor is $\Sigma = 8 + 2(\varepsilon_1 - \varepsilon_2) + 6(\varepsilon_0 - \varepsilon_3) + 2(\varepsilon_0\varepsilon_1 - 2\varepsilon_0\varepsilon_3 + \varepsilon_2\varepsilon_3)$. Importantly, the state probabilities are now quite reversible: by computing the ratio Q_{2k}/Q_{2k+1}, one can easily find all ε_k given all Q_k, and so the skew parameters can be expressed directly in terms of the triplet frequency contrasts.

IV. WEAKLY-CONSTRAINED BLOCK CODES FOR MODULATION

Initially weakly-constrained codes were presented by K.A.S.Immink in [12] and exploited in magnetic recording systems. These codes in contrast to constrained (strictly-constrained) codes do not strictly observe the channel constraints, rather their codewords violate the channel constraints with a small probability. Thanks to the method of weakly-constrained coding the most
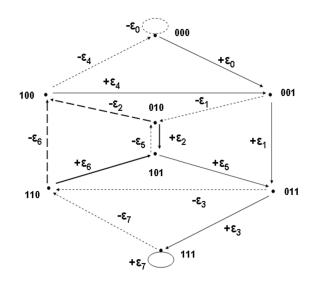

Fig. 2. The encoder model

effective RLL-codes have been constructed ([13]). Modulation codes described below are weakly-constrained also, because their application depends on the ε parameter ($0 \leq \varepsilon \leq 1$). This parameter is the "degree of weakness" for our modulation code: when $\varepsilon = 1$, the code is strongly-constrained; and when $\varepsilon = 0$ the code has no constraints whatsoever.

A. The idea of block code

The idea of modulation block code is as follows. Consider a block of m bits. Out of the 2^m bit patterns there will be some $G_{m,0}$ patterns that contain no combination 101, $G_{m,1}$ patterns that contain exactly one such combination, etc. Let $W_{m,k}$ be a list of all bit patterns (in some order) that contain no more than k 101-combinations. The list is $L_{m,k} = \sum_{i=0}^{k} G_{m,i}$ entries long. Denote as $W_{m,k}(j)$ the jth entry of the list. Now we define the block code $B_{m,k}$ thus:

$$
y_i = W_{m,k}\left(\left\lfloor \frac{s}{(L_{m,k})^i} \right\rfloor \mod L_{m,k}\right), \quad (2)
$$

where $s = \sum_{i=0}^{n-1} x_i 2^i$ is the whole n-bit source message interpreted as a binary number and each y_i is an m-bit sequence representing one output block. The output is thus the sequence $\vec{y} = (y_0, y_1, ..., y_{i_{max}})$, $0 \leq i \leq i_{max}$, where i_{max} is such that $(L_{m,k})^i \leq s < (L_{m,k})^{i+1}$. This can be thought of as the conversion of s from binary to the positional number system based $L_{m,k}$, where y_i play the role of "digits". Such a conversion is reversible. The function $W(j)$ is obviously invertible since it is an injection by construction and consequently allows unambiguous decoding.

For a long enough source sequence x_i, the code redundancy tends to

$$R = \frac{m - \log_2 L_{m,k}}{m}.$$

The per-bit density of the 101 combination in the output is approximately

$$\frac{1}{mL_{m,k}} \sum_{i=0}^{k} iG_{m,i}$$

This is an approximation that neglects the 101 combination at the junctures of neighbouring blocks. These depend on the frequency of certain bits at the beginning and end of a bit pattern. Indeed, the combination 101 will arise around the starting bit of a pattern that starts with 01, provided that it follows a pattern ending with 1. Similarly, a 101 combination around the last bit of a pattern arises when the pattern ends with 10, and the next one starts with 1. The exact per-bit frequency of the 101 combinations is thus:

$$Q_5 = \frac{1}{mL_{m,k}} \sum_{i=0}^{k} iG_{m,i} + (\phi_{e1}\phi_{b01} + \phi_{e10}\phi_{b1})/m.$$

Here $\phi_{b01}, \phi_{b1}, \phi_{e1}$, and ϕ_{e10} are frequencies of the patterns that begin with 01, begin with 1, end with 1, and end with 10, respectively, for the given m and k. Note that due to the assumed randomness of the source bit stream, the neighbouring patterns are uncorrelated, which justifies the use of products of frequencies in the above formula.

We have constructed block codes in table form for m=8,16 and 28 by exhaustively searching the bit pattern space by computer. For each code we also quantified the ϕ frequencies and computed $L_{m,k}$ for the first few values of k. For the purposes of comparing these codes with the Markov chain described earlier, we have computed the effective ε, i.e the value at which P_5 of the Markov chain matches the above value for the block code. This enabled us to plot R vs effective ε for all codes and the Markov chain in a single graph, see Fig.3.

It is clear from the figure that the Markov chain is an ideal case of the block codes, which they very slowly approximate as the code size m increases. Codes with $m = 8$ and $m = 16$ are quite practical as table codes, since the table can easily be provided in Read-Only Memory of a very small size (64K entries at most). The 28-bit code is less so as it requires 256M entries of 3.5 bytes each, totalling about 1Gb, but even a code as large as this does not approximate the redundancy level of the Markov chain well enough, especially at small ε. For larger block codes which ensure better redundancy, the table method is impractical, and an effective encoding algorithm is required. This algorithm was proposed in [14]. Using it the code with 128-bit block length has been built (its redundancy is shown in Fig.3, curve 5).

B. Theoretical limit of block code redundancy for arbitrary ε

Denote as $P(x)$ the probability that the bit sequence x occurs in the *encoded* message. For example, P(1010) is the probability that four consecutive bits taken at random will turn out to be the sequence 1010. Also denote as $S(m,k)$ a set of length-m sequences, where each sequence has *exactly* k triplets 101.

Lemma 4.1: For any $k \geq 0$, $m \geq 5$, the equation $P(00000) = P(00100)$ holds.

Proof: Since neither the replacement $00000 \rightarrow 00100$, nor its inverse modifies the frequency of 101s in the sequence, these replacements exchange a codeword $B \in S(m,k)$ with another codeword $B' \in S(m,k)$. Similarly it can be proven that $P(00001) = P(00011)$ and $P(00010) = P(00110)$. ∎

We also observe that for block codes $P_0 \neq 0$ since the 000 triplet cannot induce an occurrence of 101, so it would not be removed from any codewords. Consequently $\varepsilon_0 \neq 1$. Furthermore, $\varepsilon_0 \neq -1$, otherwise the probability of any triplet would be equal to zero except the probability of 000, which is impossible.

Bearing the above lemma in mind and remembering the transition probabilities of the Markov chain, we derive:

$P(00001) = Q_0 P_{0\to 0} P_{0\to 1}$;
$P(00011) = Q_0 P_{0\to 1} P_{1\to 3}$; hence $\varepsilon_1 = -\varepsilon_0$.
$P(00010) = Q_0 P_{0\to 1} P_{1\to 2}$ and
$P(00110) = Q_1 P_{1\to 3} P_{3\to 6}$; hence $\varepsilon_3 = \varepsilon_1 = -\varepsilon_0$.
$P(00000) = Q_0 P_{0\to 0} P_{0\to 0}$;
$P(00100) = Q_1 P_{1\to 2} P_{2\to 4}$;
hence

$$\varepsilon_2 = 1 - \frac{(1-\varepsilon_0)^3}{(1+\varepsilon_0)^2} = \frac{\varepsilon_0(\varepsilon_0^2 - 2\varepsilon_0 + 5)}{(1+\varepsilon_0)^2}.$$

The skew parameter vector for the block code is given by the following expression:

$$\vec{\varepsilon} = \varepsilon_0 \left(1, -1, \frac{\varepsilon_0^2 - 2\varepsilon_0 + 5}{(1+\varepsilon_0)^2}, -1, \right.$$
$$\left. 1, -1, \frac{\varepsilon_0^2 - 2\varepsilon_0 + 5}{(1+\varepsilon_0)^2}, -1 \right), \quad (3)$$

where every skew parameter is expressed in terms of a single skew ε_0. However, to be able to juxtapose

the multi-skew block code and a single-skew Markov chain, we must first define the basis for comparison. Such a basis could be an effective skew value ε_{eff} such that the density of 101 in the block code at a given $\vec{\varepsilon}$ corresponds to the density of 101 in the single-skew Markov chain at the skew value ε_{eff}. In other words, ε_{eff} indicates how much the single-skew Markov chain would have to be skewed to achieve the same density of the undesirable states. The latter is simply $Q_5 = (1 - \varepsilon_{\text{eff}})/(8 + 2\varepsilon_{\text{eff}})$ whereas from the Markov chain density and the $\vec{\varepsilon}$ expression above we get

$$Q_5 = \frac{\varepsilon_0^4 + 6\varepsilon_0^2 + 8\varepsilon_0 + 1}{8(2\varepsilon_0^2 + \varepsilon_0 + 1)} = (1 - \varepsilon_{\text{eff}})/(8 + 2\varepsilon_{\text{eff}}),$$

Provided that we can solve the quartic equation, this gives us ε_0 in terms of ε_{eff}.

C. Calculation of the redundancy

The above-mentioned quartic equation is of the form:

$$\varepsilon_0{}^4 + 2\varepsilon_0{}^2(3 - 8Q_5) + 8\varepsilon_0(1 - Q_5) - 8Q_5 = 0$$

and can be proven to have a single root in the interval $-1 \le \varepsilon_0 \le 1$ given that $0 \le Q_5 \le 1/8$. Obtaining the root numerically presents no technical problem, so ε_0 can be assumed to be functionally dependent on ε_{eff} in a known way.

Consequently, let us express the entropy in terms of ε_0. Using the general formula from [7] and interpreting $\vec{\varepsilon}$ as a function of ε_0 we get $h(\varepsilon_0) = -\sum_{k=0}^{7} Q_k(\vec{\varepsilon})f(\varepsilon_k(\varepsilon_0))$, where $f(x) = \frac{1+x}{2}\log_2\frac{1+x}{2} + \frac{1-x}{2}\log_2\frac{1-x}{2}$. Using Eqs. 1 and 3 we establish that

$$h(\varepsilon_0) = f(\varepsilon_0)(Q_0(\varepsilon_0) + Q_1(\varepsilon_0) + Q_3(\varepsilon_0) +$$
$$Q_4(\varepsilon_0) + Q_5(\varepsilon_0) + Q_7(\varepsilon_0)) +$$
$$f(\varepsilon_2(\varepsilon_0))(Q_2(\varepsilon_0) + Q_6(\varepsilon_0)).$$

By varying ε_{eff} (and via it, ε_0) over its range, we obtain curve 7 in figure 3. This is a theoretical curve which corresponds to block code for infinite m. Observe that block size 128 is roughly as close to the single-skew Markov chain result obtained previously as the theoretical limit of skewed block codes, i.e. curve 7, except they deviate on either side of the curve. It seems likely that for any practical purpose block size 128 will be sufficient, while any further increase in the block size would not yield a significant improvement in code redundancy, while potentially making the encoding/decoding process far more expensive. Another consideration would be the
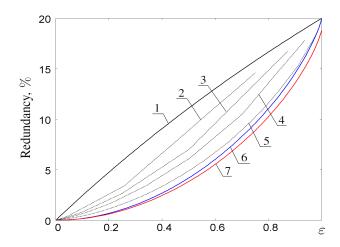

Fig. 3. $R(\varepsilon)$ graph 1 — bit-stuffing; 2 — Bk8; 3 — Bk16; 4 — Bk32; 5 — Bk128; 6 — Markov chain with one parameter ($\varepsilon_2 = -\varepsilon$); 7 — limiting (block size $\to \infty$) code redundancy.

required level of interleaving and/or FEC to mitigate significant error propagation that a block code longer than 128 would be prone to.

D. Code gain

We assume that the vector of error probabilities is, as before, $\vec{R} = K \cdot (1, 1, 1, 1, 1, M, 1, 1)$, and set $K = 10^{-3}$ for the sake of certainty. To quantify the BER improvement due to our block code we introduce an effective code gain defined as $\Gamma = 20\log_{10}(A(M, K)/A_c(M, K, \varepsilon))$, where $A(M, K)$ is the RMS of the effective AWGN over the source signal, i.e. the magnitude that corresponds to the observed BER, and where $A_c(M, K, \varepsilon)$ is the RMS of the effective AGWN after the block coding. The value of Γ corresponds to the increase of SNR that delivers the same error probability reduction in the case of AGWN without patterning as the block code delivers under the patterning defined by the vector R.

Let's present the error probability as the sum of the probabilities specified by both linear and nonlinear effects: $BER = BER_L + BER_{NL}$. $BER_L = \sum_{k=0}^{7} K \cdot Q_k = K$, $BER_{NL} = Q_5 K(M - 1)$. Since the linear and nonlinear contributions are uncorrelated, $A^2(M, K) = \sigma_L^2 + \sigma_N^2$, where σ_L^2 is the standard variation of the linear, and σ_1^2 of the nonlinear contributions. We assume that both contributions are AWGN (even though the nonlinear one is not), and so $BER_L = \frac{1}{2} \cdot \Phi\left(\frac{1}{\sigma_0\sqrt{2}}\right)$, and $BER_N = \frac{1}{2} \cdot \Phi\left(\frac{1}{\sigma_1\sqrt{2}}\right)$. Here $\Phi(x)$ is a complementary error function and BER_L and BER_N are error probabilities for the linear and nonlinear contributions, respectively. From these equations one
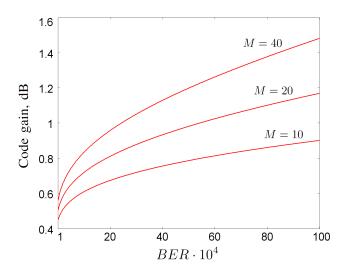
Fig. 4.   Block code gain versus BER ($R = 10.27\%$)

can obtain σ_L and σ_N for arbitrary BER values. Consequently, $A_c^2(M, K, \varepsilon) = (\sigma_L')^2 + (\sigma_N')^2$, where $(\sigma_L')^2$ is the variation of the linear contribution for the encoded signal, and $(\sigma_N')^2$ that of the nonlinear contribution. Finally we get $\Gamma = 10 \log_{10} \dfrac{\sigma_L^2 + \sigma_N^2}{(\sigma_L')^2 + (\sigma_N')^2}$.

Figure 4 shows the code gain as a function of the BER for a long block code (depicted as curve 7 in fig 3) taken at 10% redundancy. It is remarkable that a gain of about 1dB is possible from such a simple code. It should be borne in mind that the gain displayed in the figure does not take into account any processes associated with decoding (and long block codes are almost as prone to error propagation as convolution codes). Nevertheless, if adequately protected by FEC, the effects of error propagation with a moderate block size can be alleviated. It may also be possible to devise a soft decoding scheme based on the algorithm presented in the next section, which is less prone to these effects.

V. Conclusions

A detailed analysis of block codes for weakly-constrained encoding/decoding is presented following the first theoretical results in [7]. The redundancy measure is obtained against the effective factor ε for specific block codes, which in some cases turns out to be somewhat better than the theoretical model published previously. This effect is explained by the fact that the correct Markov model for real codes is multiparametric, not uniparametric as in [7]. It is remarkable, nevertheless, that the uniparametric Markov chain still approximates the real behaviour so well. The effective code gain for

practical block codes has been obtained and plotted against a range of bit error rates at various magnitudes of the patterning effect. It is shown that using these simple codes in a hard-decision situation provides an effective gain of about 1dB.

References

[1] F. Matera, A. Mecozzi, M. Settembre, I. Gabitov, H. Haunstein, and S. K. Turitsyn, "Theoretical evaluation of the noise growth and the system performance for a link constitued by a chain of N optical amplifiers with in-line filters," OFC'98 Tech. Dig., WM23, p. 202, (1998)

[2] R. J. Essiambre, B. Mikkelsen, and G. Raybon, "Intra-channel cross-phase modulation and four-wave mixing in high-speed TDM systems," *Electron. Lett.*, vol. 35, pp. 1576–1578, 1999.

[3] P. V. Mamyshev and N. A. Mamysheva, "Pulse-overlapped dispersion-managed data transmission and intrachannel four-wave mixing," *Opt. Lett.*, vol. 24, pp. 1454–1456, 1999.

[4] A. H. Gnauck, A. Mecozzi, M. Shtaif, and J. Wiesenfeld, "Modulation Scheme for Tedons," U.S. Patent Application, #20020126359, 2001.

[5] E. G. Shapiro, M. P. Fedoruk, S. K. Turitsyn, and A. Shafarenko, "Reduction of nonlinear intrachannel effects by channel asymmetry in transmission lines with strong bit overlapping," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 1473–1475, Oct. 2003.

[6] B. Vasic, V. S. Rao, I. B. Djordjevic, R. K. Kostuk, and I. Gabitov, "Ghost-pulse reduction in 40-Gb/s systems using line coding," *IEEE Photon. Technol. Lett.*, vol. 16, pp. 1784–1786, Jul. 2004.

[7] A. Shafarenko, K. S. Turitsyn, and S. K. Turitsyn, "Skewed coding for suppression of pattern-dependent errors," in *Proc. of 31st European Conference on Optical Communication (ECOC 2005)*, vol. 2, pp. 193–194, Glasgow, United Kingdom, Sept. 2005.

[8] B. Djordjevic and B. Vasic, "Nonlinear BCJR equalizer for suppression of intrachannel nonlinearities in 40 Gb/s optical communications systems," Opt. Express, vol. 14, pp. 4625–4635, May 29, 2006.

[9] S. K. Turitsyn, M. P. Fedoruk, O. V. Shtyrina, A. V. Yakasov, A. Shafarenko, S. R. Desbruslais, K. Reynolds, R. Webb, Patterning effects in a WDM RZ-DBPSK SMF/DCF optical transmission at 40Gbit/s channel rate, Optics Communications, 277(2), 264–268 (2007)

[10] Shapiro E.G., Fedoruk M.P., Turitsyn S.K., "Direct modelling of error statistics at 40 Gbit/s rate in SMF/DCF link with strong bit overlapping," Electronic Letters, vol. 40, no. 22, pp. 1436–1437, 2004.

[11] B. Slater, S. Boscolo, A. Shafarenko and S. K. Turitsyn, Mitigation of patterning effects at 40 Gbits/s by skewed channel pre-encoding, Journal of Optical Networking 6 (8), 984 (2007)

[12] K.A.S. Immink, "Weakly constrained codes", Electronic Letters, vol. 33, no. 23, pp. 1943–1944, Nov. 1997.

[13] Ming Jin, K.A.S. Immink, B. Farhang-Boroujeny, "Design Techniques for Weakly Constrained Codes", IEEE Trans. Comm., vol. 51, no. 5, pp. 709-714, May 2003.

[14] A.Shafarenko, A.Skidin, S.K.Turitsyn, "Skewed-Statistics Modulation Codes for the Mitigation of Patterning Effects in Digital Communications", IEEE Trans. Comm., to be published.