Transactions Letters

Information-Theory Analysis of Skewed Coding for Suppression of Pattern-Dependent Errors in Digital Communications

Alexander Shafarenko, *Senior Member, IEEE*, Konstantin S. Turitsyn, and Sergei K. Turitsyn

Abstract—We present information-theory analysis of the tradeoff between bit-error rate improvement and the data-rate loss using skewed channel coding to suppress pattern-dependent errors in digital communications. Without loss of generality, we apply developed general theory to the particular example of a high-speed fiber communication system with a strong patterning effect.

Index Terms—Digital communication, error correction, inter-symbol interference (ISI), information theory.

I. INTRODUCTION

PATTERNING effects due to intersymbol interference (ISI) manifest themselves in digital communication as the dependence of the transmission output for an information bit on its surrounding pattern, i.e., the neighboring bits. ISI imposes one of the most severe limitations in high-speed data transmission. Patterning effects can be due to physical mechanisms of varying nature. For instance, in fiber optic digital communication, the pattern dependence of errors can be due to the gain saturation in semiconductor optical amplifiers (see, e.g., recent papers [1] and [2]), or to resonance interactions between pulses in bit-overlapping transmission regimes [3], [4]. An important and actively studied example of transmission with pattern-dependent errors is optical fiber communication at high bit rates limited by intra-channel four-wave-mixing (ICFWM) [3], [4] through the generation of "ghost" pulses. The "ghost" pulses emerge in the time slots carrying logical "zero" bits. They are generated by resonance nonlinear ISI of periodically overlapping (due to dispersive broadening) pulses carrying logical "ones." The major contributions to the bit errors come from the ghost pulses arising in "zero" time slots surrounded by symmetric patterns of logical "ones." Various techniques have been proposed and implemented to suppress ICFWM. In this letter, using ICWFM

as a key example but without loss of generality, we consider information-theory approaches to reducing pattern-dependent errors. Suppression of nonlinear intrachannel effects by channel coding was first proposed in [5] (see also [6]). The use of the modulation codes for the prevention of ICFWM effects was proposed in [7]. In this letter, we present information-theory analysis of the tradeoff between the bit-error rate (BER) improvement and the loss of data rate using *skewed channel coding*.

Patterning effects can be partially characterized at the receiver by analysis of error rates for the eight triplets corresponding to the possible combinations of the nearest neighboring bits [8], [9]. The error probabilities for the central bit in the eight triplets can be gathered into an error vector $\overrightarrow{Q} = (Q_0, Q_1, Q_2, \ldots, Q_7)$. Total BER of a transmitted packet (neglecting the errors at its end points) is then given by the BER $= \sum_{k=0}^{7} P_k Q_k$, where P_k is the probability of the occurrence of a triplet with the index k in the input bit string, and Q_k is the error probability for the central digit in the triplet. An uneven distribution of errors Q_k over k offers an opportunity to reduce the error rate by *reducing the probability of the triplets that affect the BER most*, using skewed pre-encoding. Obviously, this can only be done at the expense of the information content of the packet, which is represented by the transmitted signal entropy H (measured in bits/digit).

II. THE SKEWED ENCODING

For illustration purposes, in the rest of the letter, we focus on the example of an ICFWM-limited system, although our approach is general. As shown in [2], the major contribution to the BER in systems limited by ICFWM is from the pattern "101." In the course of extensive numerical simulations, it was observed [2] that the probability of error for that pattern is approximately 20 times (under certain system/signal parameters) that for any other triplet. To model the error distribution between the triplets, we would like to look at the example when $Q_{101} = Q_5 = MQ_k, k \neq 5$, with the error asymmetry factor M varying from 10 to 40. Thus, we use the error vector $\overrightarrow{Q} = q(1, 1, 1, 1, 1, M, 1, 1)$. Our goal now is to quantify the BER improvement due to a pre-encoding of transmission data that makes certain combinations (in this example, 101) of the input digits less frequent.

The two most important performance characteristics of a transmission system are the BER and the channel throughput. The former is improved by pre-encoding; the latter is worsened by it. The tradeoff is consequently between the BER

Fig. 1. Transition graph.

improvement and the loss of data rate (or spectral efficiency) and increase in the complexity of an encoder/decoder. The first issue to consider is the source information content as a function of the pre-encoding skew. We use the Markov chain shown in Fig. 1 as a model of the encoder. The chain is a random process that attempts to avoid the combination 101 that was found error-prone in [2]. The vertices of the graph correspond to the state of the process, which consists of the three last digits up to and including the current one: $d_{k-2}d_{k-1}d_k$. The probability of the next digit is depicted as a transition from that state to the next one, $d_{k-1}d_kd_{k+1}$, with d_{k+1} either 1 or 0, with a probability depending on the current state $d_{k-2}d_{k-1}d_k$. We use *dashed arrows* for $d_{k+1} = 0$ and *solid* ones for $d_{k+1} = 1$. Those arrows are either thin, corresponding to the *nonskewed* transitions with the probability 0.5, or thick, corresponding to the *skewed* transitions. The latter bear the probability $(1 - \varepsilon)/2$ leading to the "bad" state 101, and the probability $(1 + \varepsilon)/2$ leading to a neighboring "good" state 100. Notice that each state whose label (read as a binary number) represents quantity k has exactly two transitions from it, into states $2k + 0 \mod 8$ and $2k + 1 \mod 8$. From information theory, the information content of a message of size n, $\overrightarrow{b}_n$, is given by its entropy

$$H_n = -\sum_{b_n} \Pr(\overrightarrow{b}_n) \log_2 \Pr(\overrightarrow{b}_n). \tag{1}$$

Here the summation is done over all 2^n possible n-bit strings.

Obviously, this works out as $H = n$ for uncorrelated messages where the probability of each string is 2^{-n}. For correlated messages, let us split the message into the prefix of size $n - 1$, $\overrightarrow{b}_{n-1}$, and the last bit d. It is easy to see that the probability

$$\Pr(\overrightarrow{b}_{n-1}, d) = \Pr(\overrightarrow{b}_{n-1})T_{b,2b+d}$$

where T is the process transition matrix, b the binary value of the string $\overrightarrow{b}_{n-1}$, and both indices of T are assumed modulo 8,

so effectively, only the three last digits of the bit string matter. Now substitute the above in (1), and assume that summation is first over $d = 0, 1$ and then all $\overrightarrow{b}_{n-1}$; also remembering that $T_{b,2b+0} + T_{b,2b+1} = 1$, we derive

$$H_n = H_{n-1} - \sum_{\overrightarrow{b}_{n-1}} \Pr(\overrightarrow{b}_{n-1})T_{b,2b+d} \log_2(T_{b,2b+d})$$

$$= H_{n-1} - \sum_{s=0}^{7} W_{n-1}(s) \sum_{d=0,1} T_{s,2s+d}$$

$$\times \log_2(T_{s,2s+d}) = H_{n-1} + h_n. \tag{2}$$

The factor $W_{n-1}(s)$ is the sum of the probabilities of all $(n - 1)$-bit strings that end in s. In other words, $W_{n-1}(s)$ is the probability that a string produced by the Markov chain has the last state s. Since the transition graph is fully connected (each state is reachable from each other state), we can assume that the Markov process is ergodic, which means that $W_{n-1}(s) \rightarrow W_n(s)$ at large n, which does not depend on n, and which, moreover, equals the density of state s in a very long individual bit string, produced by the Markov process. This density can be evaluated from the transition matrix T as follows.

Due to ergodicity, the state density $nx_n(s)$, which is the number of occurrences of state s in a string of length n produced by the chain, is a stationary random process as $n \rightarrow \infty : x_n(s) \rightarrow x(s)$. Let us introduce an eight-component vector $\overrightarrow{x} = \overrightarrow{P} = (P_0, \ldots, P_7)$ representing the stationary values $x(s)$. The stationary distribution corresponding to the Markov process presented in Fig. 1 must satisfy the condition $\overrightarrow{P}\hat{T} = \overrightarrow{P}$, where $\hat{T} = T_{k,l}$ is the process transition matrix corresponding to the graph in Fig. 1. The requirement of the stationary process yields the following system of linear equations for state populations:

$$P_0 + P_4 = 2P_0$$
$$P_0 + P_4 = 2P_1$$
$$P_1 + P_5 = 2P_2$$
$$P_1 + P_5 = 2P_3$$
$$(1 + \varepsilon)(P_2 + P_6) = 2P_4$$
$$(1 - \varepsilon)(P_2 + P_6) = 2P_5$$
$$P_3 + P_7 = 2P_6$$
$$P_3 + P_7 = 2P_7.$$

The solution of this system of equations normalized by $\sum_{k=0}^{7} P_k = 1$ gives us an eight-component vector $\overrightarrow{P}$ representing the stationary distribution of the triplet's probabilities

$$\overrightarrow{P} = \frac{1}{8 + 2\varepsilon}(1 + \varepsilon, 1 + \varepsilon, 1, 1, 1 + \varepsilon, 1 - \varepsilon, 1, 1). \tag{3}$$

Vector $\overrightarrow{P}$ gives the number of the occurrences of the triplet with the index k in an infinite string of bits. It is clear that, in particular, the frequency of occurrence is higher for the triplets 000, 001, and 100, and lower for 101. Now remember that the components of the vector $\overrightarrow{P}$ are exactly the values of $W(s)$
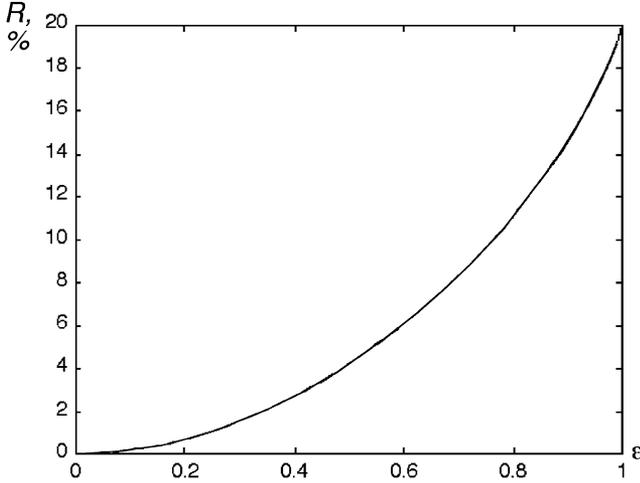
Fig. 2.   Redundancy R, %, versus skew parameter ε.

Fig. 3.   Code gain ΔG versus redundancy R, %, added by skewed coding.

referred to above. Coming back to (2), we are ready to quantify the per-bit entropy $h = \lim_{n\to\infty} h_n$ which is given by

$$h = -\sum_{k=0}^{7} P_k(T_{k,2k}\log_2 T_{k,2k} + T_{k,2k+1}\log_2 T_{k,2k+1}). \quad (4)$$

III. CALCULATION OF THE REDUNDANCY

Redundancy of the signal coding with nonuniform probabilities of different patterns is defined as $R = h_{\max} - h = 1 - h$. Substituting the stationary distribution $\overrightarrow{P}$ into (4) yields the per-bit entropy $h = (3 + g(\varepsilon) + \varepsilon)/(4 + \varepsilon)$, and the redundancy $R = (1 - g(\varepsilon))/(4 + \varepsilon)$, where

$$g(\varepsilon) = -\frac{(1+\varepsilon)}{2}\log_2\frac{(1+\varepsilon)}{2} - \frac{(1-\varepsilon)}{2}\log_2\frac{(1-\varepsilon)}{2}.$$

We are now in a position to quantify the tradeoff between the improvement of the error rate and the loss of the information content (which is the effective throughput of the transmission system) using skewed pre-encoding.

Fig. 2 displays the result in graphic form. The largest redundancy corresponds to $\varepsilon = 1$ (extreme skew), where 20% of the throughput is lost. However, under a significant skew of 25%, the loss of throughput is only 1%, which suggests that skewed coding could be bandwidth-efficient when used in addition to standard forward-error correcting (FEC) codes. To quantify the BER improvement due to the skewed code, we introduce a code gain factor, defined as

$$\Delta G = 10\log_{10}[\text{BER}(\varepsilon = 0)/\text{BER}(\varepsilon)].$$

Note that the term "code gain" is used in a different context compared with the standard FEC notation (which involves the energy per bit parameter) commonly used to describe linear additive white Gaussian noise (AWGN) channels.

Fig. 3 shows the BER improvement as a function of code redundancy. We observe that reasonable improvement can be achieved with a relatively small redundancy. Evidently, the skewed coding becomes more efficient with the increase of the

Fig. 4.   Code gain versus error asymmetry factor M.

parameter M. Fig. 4 similarly presents code gain versus error asymmetry factor M for different values of the skew parameter.

IV. GENERAL CASE OF SKEWED CODING

In this section, we present the results of calculation of the redundancy due to skewed coding in the general case of an arbitrary imbalance in the probabilities of elementary transitions between the triplets. The population of the Markov states in the general case (illustrated in Fig. 5) can be determined by introducing the offsets between state transitions as follows:

$$000 \to 000 = (1 - \varepsilon_0)/2$$
$$000 \to 001 = (1 + \varepsilon_0)/2$$
$$001 \to 010 = (1 - \varepsilon_1)/2$$
$$001 \to 011 = (1 + \varepsilon_1)/2$$
$$010 \to 100 = (1 - \varepsilon_2)/2$$
$$010 \to 101 = (1 + \varepsilon_2)/2$$
$$011 \to 110 = (1 - \varepsilon_3)/2$$
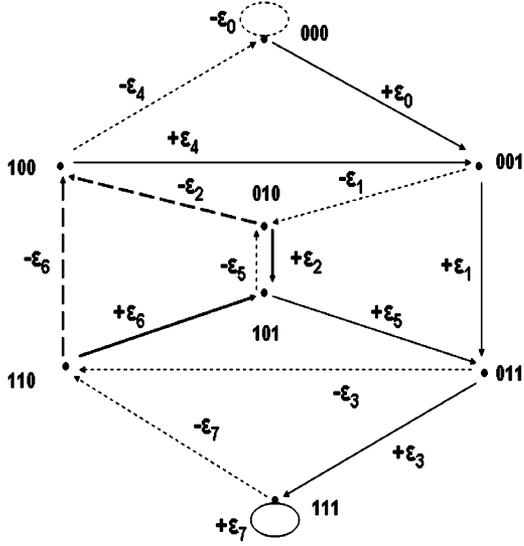
Fig. 5.   Transition graph in the general case.

$$011 \rightarrow 111 = (1 + \varepsilon_3)/2$$
$$100 \rightarrow 000 = (1 - \varepsilon_4)/2$$
$$100 \rightarrow 001 = (1 + \varepsilon_4)/2$$
$$101 \rightarrow 010 = (1 - \varepsilon_5)/2$$
$$101 \rightarrow 011 = (1 + \varepsilon_5)/2$$
$$110 \rightarrow 100 = (1 - \varepsilon_6)/2$$
$$110 \rightarrow 101 = (1 + \varepsilon_6)/2$$
$$111 \rightarrow 110 = (1 - \varepsilon_7)/2$$
$$111 \rightarrow 111 = (1 + \varepsilon_7)/2.$$

It is easy to observe the following relations between probabilities of different triplets:

$$P_{001} \equiv P_1 = P_4 \equiv P_{100}$$
$$P_{110} \equiv P_6 = P_3 \equiv P_{011}$$
$$P_1 = \frac{1 + \varepsilon_0}{1 - \varepsilon_4} P_0$$
$$P_6 = \frac{1 - \varepsilon_7}{1 + \varepsilon_3} P_7.$$

After straightforward algebra, the solution of the general imbalanced stationary problem $\overrightarrow{P}\hat{T} = \overrightarrow{P}$ is found as

$$P_0 = \frac{1}{M_1 + M_2 K}, \quad P_1 = P_4 = \frac{(1 + \varepsilon_0)}{(1 - \varepsilon_4)} P_0,$$

$$P_7 = K P_0, \quad P_3 = P_6 = \frac{(1 - \varepsilon_7)}{(1 + \varepsilon_3)} K P_0,$$

$$P_2 = \left[\frac{(\varepsilon_5 - \varepsilon_1)}{(1 + \varepsilon_5)} \frac{(1 + \varepsilon_0)}{(1 - \varepsilon_4)} + \frac{(1 - \varepsilon_5)}{(1 + \varepsilon_5)} \frac{(1 - \varepsilon_7)}{(1 + \varepsilon_3)} K \right] P_0,$$

$$P_5 = \left[\frac{(1 + \varepsilon_2)}{(1 - \varepsilon_2)} \frac{(1 + \varepsilon_0)}{(1 - \varepsilon_4)} + \frac{(\varepsilon_6 - \varepsilon_2)}{(1 - \varepsilon_2)} \frac{(1 - \varepsilon_7)}{(1 + \varepsilon_3)} K \right] P_0$$

where

$$M_1 = 1 + 2 \frac{(1 + \varepsilon_0)}{1 - \varepsilon_4} + \frac{(1 + \varepsilon_2)}{(1 - \varepsilon_2)} \frac{(1 + \varepsilon_0)}{(1 - \varepsilon_4)}$$
$$+ \frac{(\varepsilon_5 - \varepsilon_1)}{(1 + \varepsilon_5)} \frac{(1 + \varepsilon_0)}{(1 - \varepsilon_4)}$$

$$M_2 = 1 + 2 \frac{(1 - \varepsilon_7)}{1 + \varepsilon_3} + \frac{(1 - \varepsilon_5)}{(1 + \varepsilon_5)} \frac{(1 - \varepsilon_7)}{(1 + \varepsilon_3)}$$
$$+ \frac{(\varepsilon_6 - \varepsilon_2)}{(1 - \varepsilon_2)} \frac{(1 - \varepsilon_7)}{(1 + \varepsilon_3)}$$

$$K = \frac{[2 + \varepsilon_1 + \varepsilon_5 + \varepsilon_2(\varepsilon_5 - \varepsilon_1)]}{[2 - \varepsilon_2 - \varepsilon_6 + \varepsilon_5(\varepsilon_2 - \varepsilon_6)]} \frac{(1 + \varepsilon_0)}{(1 - \varepsilon_4)} \frac{(1 + \varepsilon_3)}{(1 - \varepsilon_7)}.$$

This general solution can be further simplified by making some assumptions regarding the encoding procedure. These additional assumptions can be attributed to the properties of the codes used to generate a bit stream with the required information features. For instance, assuming that no more than degree-2 neighbors have impact on the bit coding, we obtain additional symmetries for state transitions

$$100 \rightarrow 000 = 000 \rightarrow 000$$
$$100 \rightarrow 001 = 000 \rightarrow 001$$
$$101 \rightarrow 010 = 001 \rightarrow 010$$
$$101 \rightarrow 011 = 001 \rightarrow 011$$
$$010 \rightarrow 100 = 110 \rightarrow 100$$
$$110 \rightarrow 101 = 010 \rightarrow 101$$
$$011 \rightarrow 110 = 111 \rightarrow 110$$
$$011 \rightarrow 111 = 111 \rightarrow 111.$$

Applying these symmetry relations, the general solution given above, is reduced to the following expressions for the probabilities of the state (triplet) populations: $\overrightarrow{P} = (P_0, \ldots, P_7)$

$$P_0 = \frac{(1 - \varepsilon_0)(1 - \varepsilon_2)(1 - \varepsilon_3)}{\Sigma}$$
$$P_1 = P_4 = \frac{(1 + \varepsilon_0)(1 - \varepsilon_2)(1 - \varepsilon_3)}{\Sigma}$$
$$P_2 = \frac{(1 + \varepsilon_0)(1 - \varepsilon_1)(1 - \varepsilon_3)}{\Sigma}$$
$$P_3 = P_6 = \frac{(1 + \varepsilon_0)(1 + \varepsilon_1)(1 - \varepsilon_3)}{\Sigma}$$
$$P_5 = \frac{(1 + \varepsilon_0)(1 + \varepsilon_2)(1 - \varepsilon_3)}{\Sigma}$$
$$P_7 = \frac{(1 + \varepsilon_0)(1 + \varepsilon_3)(1 + \varepsilon_1)}{\Sigma}$$

where the normalizing factor is

$$\Sigma = 8 + 2(\varepsilon_1 - \varepsilon_2) + 6(\varepsilon_0 - \varepsilon_3) + 2(\varepsilon_0 \varepsilon_1 - 2\varepsilon_0 \varepsilon_3 + \varepsilon_2 \varepsilon_3).$$

V. DISCUSSION

The analysis technique described in this paper can also be applied to areas such as magnetic recording, where modulation codes are routinely used for the avoidance of undesirable bit patterns. The most common modulation codes in these applications are $RLL(n, m)$ codes, and more recently, maximum transition run (MTR) codes. The former eliminate the runs of zeros of length less than n or more than m. The latter restrict the 0-1 and 1-0 transition frequency [10]. Paper [11] is often cited for its presentation of the principles and information-theoretical techniques relevant to the analysis of such types of modulation. It defines a method of calculating the redundancy of specific codes that can be defined in graph form, but it does not

introduce pattern-resistance properties as a function of redundancy for a general modulation code. For that latter purpose, our analysis based on a Markov chain appears uniquely suitable. The direct application of our results to graph codes for recording media is easily achieved by making certain transition probabilities of the Markov chain extremely skewed ($\varepsilon = \pm 1$), while keeping others at zero skew. Our method will immediately compute the capacity of a small enough graph code (and then a larger code would merely necessitate a larger Markov chain). However, our method has exactly the same complexity for analyzing soft probabilistic constraints ($0 < |\varepsilon| < 1$), which recording media applications do not employ yet; our analysis shows that such constraints are useful in fiber optic communications. Future research may show their utility for modulation coding elsewhere.

VI. CONCLUSION

We have presented an information-theory approach to the improvement of BER in systems degraded by pattern-dependent errors. Decrease of the error rate is achieved by application of a skewed channel coding that reduces the probability of occurrence for the triplets that make the main contribution to the BER. As a particular example, we applied the theory to an important and actively studied example of transmission with pattern-dependent errors, namely, high-speed (more than 40 Gb/s/channel) optical fiber communication limited by pattern-dependent ICFWM. We have quantified the tradeoff between the BER improvement and the data-rate loss due to skewed coding.

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.
[2] E. G. Shapiro, E. G. Shapiro, M. P. Fedoruk, and S. K. Turitsyn, "Direct modelling of error statistics at 40 Gbit/s rate in SMF/DCF link with strong bit overlapping," *Electron. Lett.*, vol. 40, no. 22, pp. 1436–1437, 2004.
[3] R. J. Essiambre, B. Mikkelsen, and G. Raybon, "Intrachannel cross phase modulation and four wave mixing in high speed TDM systems," *Electron. Lett.*, vol. 35, pp. 1576–1578, 1999.
[4] P. V. Mamyshev and N. A. Mamysheva, "Pulseoverlapped dispersion-managed data transmission and intrachannel four-wave mixing," *Opt. Lett.*, vol. 24, pp. 1454–1456, 1999.
[5] A. H. Gnauck, A. Mecozzi, M. Shtaif, and J. Wiesenfeld, "Modulation Scheme for Tedons," U.S. Patent Application, #20020126359, 2001.
[6] 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, no. 10, pp. 1473–1475, Oct. 2003.
[7] 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, no. 7, pp. 1784–1786, Jul. 2004.
[8] E. Iannone, F. Matera, A. Mecozzi, and M. Settembre, *Nonlinear Optical Communication Networks*.   New York: Wiley, 1998.
[9] C. J. Anderson and J. A. Lyle, "Technique for evaluating system performance using Q in numerical simulations exhibiting intersymbol interference," *Electron. Lett.*, vol. 30, pp. 71–72, 1994.
[10] J. Moon and B. Brickner, "Maximum transition run codes for data storage systems," *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 3992–3994, Sep. 1996.
[11] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 1, pp. 5–37, Jan. 1992.