An Improved Pragmatic Turbo Encoding Scheme for High Spectral Efficiency Using Constellation Shaping

Dan Raphaeli, *Senior Member, IEEE* and Assaf Gurevitz

***Abstract*—We propose a new turbo encoding scheme for high spectral efficiency with performance close to the Gaussian channel capacity. The scheme combines nonuniform signaling on a Gaussian channel, and pragmatic turbo coded modulation for simple and flexible implementation. A table is used to map equiprobable input symbols into non equiprobable points in the QAM constellation. It is shown that the new scheme provides shaping gains of 0.6 dB and 0.93 dB, at rates 2 and 3 bits/dim respectively compared to the equiprobable pragmatic turbo coded modulation, and reach about 1 dB from the Gaussian channel capacity.**

I. Introduction

Constellation shaping can provide an energy saving called shaping gain in addition to the usual coding gain. The idea behind constellation shaping is that signals with large norm are used less frequently than signals with small norm, thus improving the overall gain by adding shaping gain to their original coding gain [1]. The nonuniform signaling reduces the entropy of the transmitter output, and hence the average bit rate. However, if points with small energy are chosen more often than points with large one, energy savings may compensate for this loss in bit rate. Theoretically, when constellation points are selected according to a continuous Gaussian distribution at every dimension, the maximum shaping gain can be achieved. Practically, a smaller gain can be achieved in finite constellations.

Various methods have been proposed for nonuniform signaling on Gaussian channels. Gallager [2] showed that binary codes can be used for assignment of nonequiprobable discrete distributions that achieve capacity. Calderbank and Ozarow [3] introduced subconstellation partitioning. In this approach a signal constellation is partitioned into several subconstellations, points in the same subconstellation are used equiprobably, and a shaping code selects the subconstellations with various frequencies. Kschischang and Pasupathy [4] took another approach by mapping simple variable-length prefix codes onto the constellation. Using the Huffman procedure, prefix codes that approach the optimal performances were designed.

For Gaussian channels, turbo coded modulation techniques can be broadly classified into binary schemes and Turbo Trellis Coded Modulation (TTCM) [5]. The first group can be further divided into “pragmatic” schemes with a single component binary turbo code, and multilevel binary turbo codes [6]. The pragmatic approach for a bandwidth efficient turbo coding scheme has been presented in [7]. This approach is simple and versatile and is much less complex to design and to implement than TTCM. It uses only one turbo decoder, and by modifying its puncturing function and modulation signal constellation, it can obtain a large family of turbo coded modulation schemes. It is possible to include shaping in the framework of multi-level codes [6]. The multilevel approach may be less attractive than the pragmatic approach due to its increased complexity, delay (from the multistage decoding) and sophisticated design rules. The shaping that was applied to multi-level codes is multi dimensional trellis shaping code which is much more complex than the method proposed here. The performance obtained in the multilevel approach is similar (within 0.1dB) to our results. TTCM was shown to obtain the highest performance in the coding of equiprobable QAM. Fragouli and Wessel [8] have shown that by careful code selection and interleaver design it is possible to reach SNR of 0.5-0.6 dB from the (constrained) capacity. Still, the shaping gain obtained in this paper for large number of bits/dim exceeds the gain of about 0.4dB of TTCM over pragmatic approach.

We have not tried to fully optimize the interleaver or the encoder, but rather to present the general idea. It is believed that by using improved binary schemes as presented by Benedetto and Montorsi [9], a gain of up to 0.2dB more than the results presented here is possible.

In this paper we show how to create non uniform signaling in the pragmatic binary turbo coded modulation environment. Our method can be easily applied to any binary turbo or turbo-like code, including parallel, serial and Low Density Parity Check (LDPC) codes. Such results will be the subject of future research. Our nonuniform signaling scheme applies a table that maps b–bits equiprobable input words into nonequiprobable m–bits M–ary PAM (or equivalently M^2 QAM) symbols, at rates 2 and 3 bits/dim. For further improvement we also apply a feedback to the soft output mapper as suggested in [10]. For simplicity, we use a one-dimensional PAM constellation, which is equivalent to a square QAM constellation.

II. Signal Shaping in a finite constellation

It is well known that signal shaping provides further gain by replacing a uniform signal distribution by a Gaussian-like distribution in order to reduce average transmit power. As shown in [1], the maximum gain due to *shaping* can be achieved only for large constellations (with an infinite number of signals) using the continuous approximation. This leads to the conclusion that the maximum shaping gain, or the maximum reduction in average transmit power for rates

R bits/dim $\rightarrow \infty$ is given by

$$G_{s,\max} = \frac{\pi e}{6} = 1.53 \text{ dB} \tag{1}$$

But in practice, the input distribution is not continuous, and when using signal sets with finite transmission rates, this gain can never be achieved. Therefore, in order to calculate the real gain that can be achieved when using small signal sets, we turn to the calculation of the capacity gain, which is the optimization of the mutual information of the additive white Gaussian noise channel (AWGN) with discrete inputs. Consider an AWGN channel having discrete inputs $\mathbf{c}$ taking on the values $\{c_j\}$ for $j = 0 \dots J-1$, with probabilites $\mathbf{P}_r = \{P_r(c_0), P_r(c_1), \dots, P_r(c_{J-1})\}$. Denoting the noise variance by σ^2, we can express the output probability density functions as

$$Q(Y|c_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(Y-c_j)^2} \tag{2}$$

The capacity of the discrete input channel is given by the maximum of the mutual information

$$\begin{aligned} C &= \max_{\mathbf{P}_r} I(\mathbf{c};Y) \\ &= \max_{\mathbf{P}_r} \sum_{j=0}^{J-1} P_r(c_j) \int_{-\infty}^{\infty} Q(Y|c_j) \log \frac{Q(Y|c_j)}{\sum_i P_r(c_i) Q(Y|c_i)} dy \end{aligned} \tag{3}$$

We also assume a constraint on the maximum input signal power S

$$\sum_{j=0}^{J-1} P_r(c_j) c_j^2 \leq S \tag{4}$$

and normalize $\mathbf{c}$ such that

$$\frac{S}{\sigma^2} = \frac{S}{N} = \sum_{j=0}^{J-1} \frac{1}{J} c_j^2 \tag{5}$$

Optimizing the mutual information with respect to the input probabilities will give the lowest average input power when transmitting at various rates $R = C$ bits/dim. We consider the power reduction, compared to equiprobable transmission, as the desired capacity gain. The optimization procedure for all possible input probabilities is difficult. Therefore we force the channel input to be (discrete) Gaussian distributed

$$P_r(c_j) = K(\lambda) \cdot e^{-\lambda |c_j|^2}, \qquad \lambda \geq 0 \tag{6}$$

where

$$K(\lambda) = \big(\sum_{c_j} e^{-\lambda |c_j|^2}\big)^{-1} \tag{7}$$

is the distribution normalization factor. The parameter λ governs the tradeoff between average power σ_c^2 of signal points and entropy $H(\mathbf{c})$. For $\lambda = 0$, the uniform distribution arises, while increasing λ results in more concentrated distributions close to the origin. The idea is best explained using an example.

Example 1: Consider the transmission of $R = 2.0$ bits/dimension using a 16-PAM signal constellation. In this constellation the signal set $\mathbf{c}$ is one of the one-dimensional signals $-15, -13, -11, \dots, -1, 1, 3, 5, \dots, 13, 15$. We first assume the uniform distribution for each constellation point of $P_r(c_j) = 1/16$. Applying (3) (4) and (5), we get the average input power $S = 17.9$ for rate $R = 2.0$ bits/dimension. If we apply discrete Gaussian distribution (6) with an optimized value of $\lambda = 1/34$ to the constellation points $\mathbf{c}$ and use (3) again, we get the average power $S = 15.0$. Therefore the capacity gain that we achieve here is

$$10 \log \frac{17.9}{15.0} = 0.768 \text{ dB} \tag{8}$$

Similarly, using the same equations for rate $R = 3.0$ bits/dimension with $\lambda = 1/84$, we achieve capacity gain

$$10 \log \frac{82.6}{64.5} = 1.074 \text{ dB} \tag{9}$$

The losses in (8) and (9) with respect to the gains achieved by the ideal continuous AWGN channel at similar rates are 0.001 dB and 0.106 dB respectively.

III. Combining Shaping and the Pragmatic Binary Turbo Coded Modulation

A. Construction of a binary distribution.

We apply the theoretical considerations of the previous section to practical schemes. A shaping algorithm has to generate a distribution of signal points approximating the discrete Gaussian distribution (6). We propose a simple way to generate unequal probabilities that are not discrete Gaussian but produce capacity gains that are fairly close to the upper bounds achieved in (8) and (9).

Consider the binary probabilities 2^{-k} where $k = 3, 4, 5 \dots$, as was first suggested in [2], we apply these probabilities to the 16-ary PAM constellation in the following way:

$$\begin{aligned} P_r(c = \pm 1, \pm 3) &= 2^{-3}, \\ P_r(c = \pm 5, \pm 7, \pm 9) &= 2^{-4}, \\ P_r(c = \pm 11) &= 2^{-5}, \\ P_r(c = \pm 13, \pm 15) &= 2^{-6} \end{aligned} \tag{10}$$

If we use this probability mapping again in (3) for transmission rates $R = 2.0$ and 3.0 bits/dim, we achieve gains

TABLE I

SIGNAL MAPPER WITH BINARY PROBABILITIES.

Signal Point	b_0	b_1	b_2	b_3	b_4	b_5
15	0	0	0	0	0	0
13	0	0	0	0	0	1
11	0	0	0	0	1	$\times 2$
9	0	0	0	1	$\times 2$	$\times 2$
7	0	0	1	1	$\times 2$	$\times 2$
5	0	0	1	0	$\times 2$	$\times 2$
3	0	1	1	$\times 2$	$\times 2$	$\times 2$
1	0	1	0	$\times 2$	$\times 2$	$\times 2$
-1	1	1	0	$\times 2$	$\times 2$	$\times 2$
-3	1	1	1	$\times 2$	$\times 2$	$\times 2$
-5	1	0	1	0	$\times 2$	$\times 2$
-7	1	0	1	1	$\times 2$	$\times 2$
-9	1	0	0	1	$\times 2$	$\times 2$
-11	1	0	0	0	1	$\times 2$
-13	1	0	0	0	0	1
-15	1	0	0	0	0	0

of 0.682 dB and 0.948 dB, which are quite close to the capacity gains of the discrete Gaussian distribution achieved in (8) and (9). Moreover, these probabilities can easily be implemented by using a table that maps equiprobable input words of 6 bits into nonequiprobable words of 4 bits having the probabilities specified in (10). Table I shows a way to do it. The columns $b_0, b_1 \ldots b_5$ represent the input bits. The signal points in the first column show how to map each input word into one of the 16 signals. The notation [$\times 2$] in the table means that a certain input bit in that location can take on the values 0 or 1. In this way four words are mapped, for example, to the signal point 9.0 and eight are mapped to 1.0 and so on. Clearly, the probability of each input word is $1/64$, whereas probabilities of the output words become $8/64$, $4/64$, $2/64$ and $1/64$. As requested, the output probabilities are equal to the binary probabilities in (10).

It should also be noted that the table preserves "Gray" mapping for the three MSB bits, which is crucial for the performance of pragmatic binary turbo coded modulation decoding.

B. Applying to Turbo Coded Modulation

We now apply the nonequiprobable distribution derived above to Turbo Coded Modulation. In pragmatic binary turbo coded modulation [7] a single binary turbo code of rate $1/3$ is used as the component code. In this paper we will assume that the rate 1/3 code of Berrou [13] is used. Its encoder outputs are suitably multiplexed and punctured to obtain $\tilde{m}$ parity bits and $m - \tilde{m}$ information bits, as shown in Fig.1. The encoded bits are mapped into an M-PSK or M-QAM signal set. For simplicity, we used an M-PAM signal set, which is equivalent to an M^2-QAM having a spectral efficiency of $2(m - \tilde{m})$ bits/s/Hz. In the following, we will describe the main blocks of the transmitter and receiver along with design issues related to the both for equiprobable and nonequiprobable schemes.

C. The Puncturer

The puncturer block determines the code rate r_c of the turbo encoder. Let N be the number of information bits in the input block of the encoder and $L = 3N$ be the number of bits contained in the block output of the encoder. We use a simple puncturing pattern that removes arbitrarily some of the parity-check bits in each encoded block. Selection of puncturing pattern does not have a great effect on the performance of high rate turbo-codes [11]. However, we leave the information bits unpunctured to obtain better results from the iterative decoding process. After puncturing, we get an output block length $L_{punc} = N/r_c$. The flexibility of the system is achieved by the ability to use various puncturing rates combined with various mapping and modulation schemes.

D. The Interleaver

The bit interleaver that follows the puncturer block is constructed of m different random interleavers that are applied in parallel to each of the bits and preserve their order in the symbol. Each interleaver's size is $N/(r_c \cdot m)$, where N is the block length, and r_c is the code rate (after puncturing). The aim of the interleaver is to spread as much as possible, after deinterleaving, the bits associated to the same channel symbol.

E. Signal Mapper

The signal mapper associates each word of *m* encoded bits to one of the M-PAM channel symbols available in the modulator. Mapping is performed differently with respect to the signaling method. If an equiprobable signaling scheme is used, we map *m* encoded bits into one of

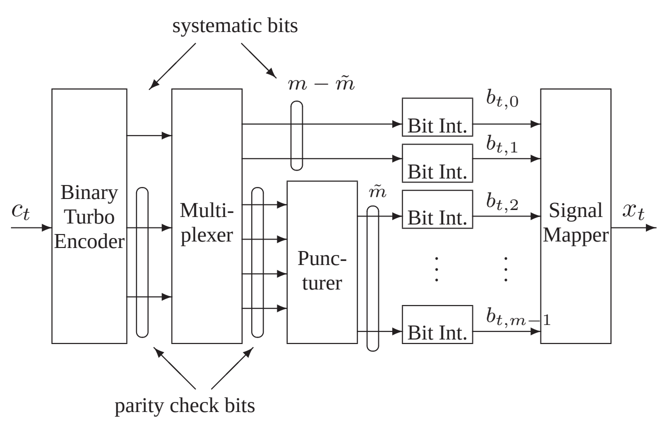

Fig. 1. A pragmatic binary turbo coded modulation encoder

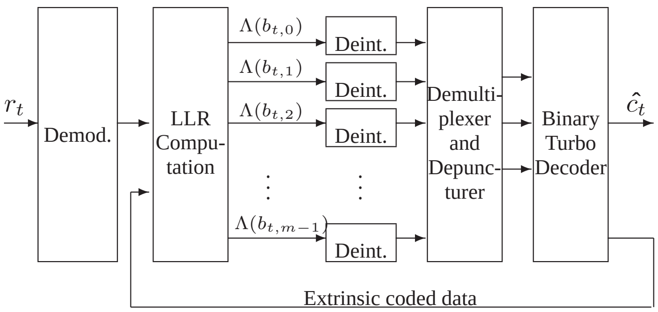

Fig. 2. A pragmatic binary turbo coded modulation decoder

the $M = 2^m$ symbols using Gray code. Otherwise, in a nonequiprobable scheme we apply a table (e.g., Table I) that maps m-bits equiprobable input words into nonequiprobable M-ary PAM symbols.

F. The Receiver

The receiver, shown in Fig.2, calculates the log-likelihood function (Bit LLR Computation block) for each encoded binary digit based on the received noisy symbol and the signal subsets in the signal constellation specified by each binary digit. The stream of the bit likelihood values is then bit deinterleaved, demultiplexed and depunctured before passing to the binary turbo decoder which is based on MAP algorithm e.g. [12]. We will describe in the following the operation of the bit LLR computation block. Continuing *Example 1*, we map each six binary encoded digits into the 16-PAM constellation using Table I for transmission rate of 4 bits/s/Hz. Let us denote the six encoded bits at time instant t by $b_{t,i}, i = 0,1,2,...,5$. According to the table, these digits select one of the 16-ary signals, which we denote by x_t. Assuming perfect coherent detection and no ISI, the received signal at instant t, denoted by r_t, can be represented by

$$r_t = x_t + n_t \tag{11}$$

where n_t is Gaussian noise with zero mean and variance σ^2. The log-likelihood ratio (LLR) for each encoded bit $b_{t,i}$, denoted by $\Lambda(b_{t,i})$, is

$$\Lambda(b_{t,i}) = \log \frac{P_r\{b_{t,i} = 1|r_t\}}{P_r\{b_{t,i} = 0|r_t\}} \tag{12}$$

We can compute the APP's of (12) Using Bayes' rule

$$\begin{aligned} P_r\{b_{t,i} = 1|r_t\} &= \frac{\sum_{x:b_{t,i}=1} P_r(b_{t,i} = 1, x, r_t)}{P_r(r_t)} \qquad (13) \\ &= \frac{\sum_{x:b_{t,i}=1} P_r(r_t|x) \cdot P_r(x|b_{t,i} = 1)}{P_r(r_t)} \end{aligned}$$

where x is the set of the signal mapper outputs, having inputs $b_{t,i} = 0$ or 1. The log-likelihood ratio $\Lambda(b_{t,i})$ is then

$$\Lambda(b_{t,i}) = \log \frac{\sum\limits_{x:b_{t,i}=1} exp(-\frac{1}{2\sigma^2}(r_t - x)^2) \cdot P_r(x|b_{t,i} = 1)}{\sum\limits_{x:b_{t,i}=0} exp(-\frac{1}{2\sigma^2}(r_t - x)^2) \cdot P_r(x|b_{t,i} = 0)} \tag{14}$$

There is a difference whether we use equiprobable or nonequiprobable signaling. In equiprobable signaling, the factor $P_r(x|b_{t,i} = 1, 0)$ in (14) cancels out because given a particular $b_{t,i}$ all the $x's$ share the same probability of $2^{-(m-1)} \ldots$ or 0.

On the other hand, in nonequiprobable signaling the probability of each x is different when $b_{t,i}$ is known. For instance, let us assume $b_{t,0} = 0$, then for equal *a-priori* probabilities of $b_{t,i}, i \neq 0$ we get by applying Table I

$$\begin{aligned} P_r(x = 15, 13|b_{t,0} = 0) &= 2^{-5}, \\ P_r(x = 11|b_{t,0} = 0) &= 2^{-4}, \\ P_r(x = 9, 7, 5|b_{t,0} = 0) &= 2^{-3}, \\ P_r(x = 3, 1|b_{t,0} = 0) &= 2^{-2}, \\ P_r(x = -1, \ldots, -15|b_{t,0} = 0) &= 0. \end{aligned} \tag{15}$$

G. Use of the LLR calculation block into the iterative decoding process

As suggested in [10], we use the bit LLR calculation block (denoted as *Soft Output Mapper*) in the turbo decoder iterations. The extrinsic coded data shown in Fig. 2 are the added values of the output turbo decoder MAP LLR's, for both information and code bits, in each turbo iteration. The data will be used by the soft output mapper as *a priori* input to the next iteration. We can write for the *a priori* probabilities in the soft output mapper,

$$P_r(b_{t,i} = 1) = \frac{e^{\Lambda_e(b_{t,i})}}{1 + e^{\Lambda_e(b_{t,i})}} \tag{16}$$

and

$$P_r(b_{t,i} = 0) = 1 - P_r(b_{t,i} = 1) \tag{17}$$

where $\Lambda_e(b_{t,i})$ is the extrinsic information for each bit. The calculation in (15) can now use the *a priori* probabilities in (16) and (17) as follows

$$P_r(x|b_{t,i}) = \prod_{k=0,k\neq i}^{m-1} P_r(b_{t,k}) \tag{18}$$

where $P_r(b_{t,k})$ are the *a priori* probabilities of $b_{t,k}$ according to the mapping of x. The calculation of (15) now becomes (few examples given),

$$
\begin{aligned}
P_r(x=15|b_{t,0}=0) &= \prod_{k=1}^{5} P_r(b_{t,k}=0),\\
P_r(x=11|b_{t,0}=0) &= \prod_{k=1}^{3} P_r(b_{t,k}=0)\cdot P_r(b_{t,4}=1),\\
P_r(x=9|b_{t,0}=0) &= \prod_{k=1}^{2} P_r(b_{t,k}=0)\cdot P_r(b_{t,3}=1),\\
P_r(x=-1|b_{t,0}=0) &= 0
\end{aligned} \tag{19}
$$

IV. Simulation results

We used the standard turbo-encoder [13], made up of two elementary encoders with the same constraint length $K=5$ and the same polynomial generators (23,35). Turbo decoding was performed in 18 iterations on blocks of 32,768 bits using random interleaving. We applied two schemes for spectral efficiencies of 2 and 3 bits/dim. The first one used a rate 1/3 turbo encoder and nonequiprobable signaling. This scheme applied Table I for a mapping of 6 bits input words including 2 information bits into 16-PAM symbols having a binary distribution as in (10). The scheme was compared to a rate 2/3 turbo code using equiprobable signaling with 3 bits 8-PAM symbols, as in the standard pragmatic binary turbo coded modulation technique. The second scheme used a rate 1/2 turbo encoder and a similar non equiprobable signaling technique, where in this time 3 information bits were mapped into 6 bits input words. It was compared with a rate 3/4 equiprobable 16-PAM pragmatic binary turbo coded modulation. Both schemes used a bit interleaver between encoder and mapper, and applied the turbo decoder extrinsic information to the soft output signal mapper. The BER versus Eb/N_0 for the two schemes are shown in Fig. 3 and 4. We can notice that for rate of 2.0 bits/dim, the nonequiprobable scheme produces a gain of 0.6 dB compared to the equiprobable one. The only power constrained, continuous input channel capacity limit of the AWGN channel is 5.74 dB. The performance of our decoder at $P_b(e)=10^{-5}$ is about 1.1 dB from this limit. At 3.0 bits/dim, the nonequiprobable scheme produces a gain of 0.93 dB compared to the equiprobable one, the channel capacity limit is 10.2 dB, and we achieve $P_b(e)=10^{-5}$ within 1.2 dB of this limit.

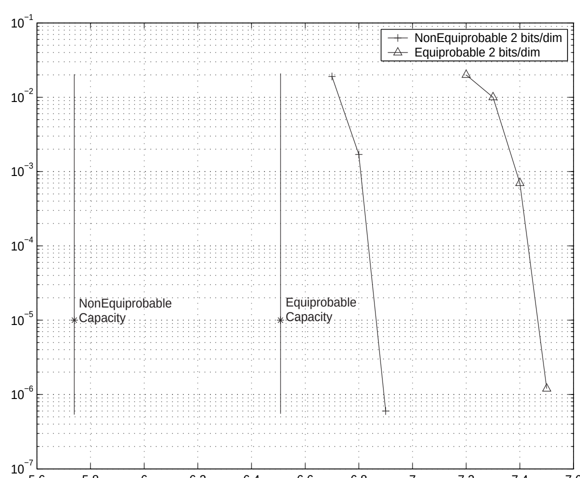

Fig. 3. Performance comparison between two schemes of Nonequiprobable and equiprobable signaling at rate 2 bits/dim using pragmatic binary turbo coded modulation with 18 iterations and Block length $N=32,768$ bits. channel capacity limit is 5.74 dB.

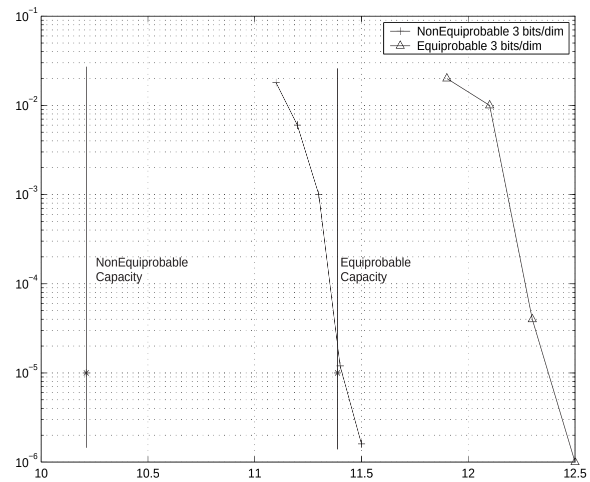

Fig. 4. Performance comparison between two schemes of Nonequiprobable and equiprobable signaling at rate 3 bits/dim using pragmatic binary turbo coded modulation with 18 iterations and Block length $N=32,768$ bits. channel capacity limit is 10.2 dB.

V. Conclusions

In this paper we presented a new scheme for improving the performance of pragmatic binary turbo coded modulation by using nonequiprobable signaling. We described a nonequiprobable signaling technique that makes it possible to approach the maximum capacity gain of a finite constellation AWGN channel. Our nonuniform signaling scheme is very easy to implement and adds negligible load on the turbo decoder. We showed for an example of 6 bits/QAM symbol, a gain of 0.93 dB out of the available shaping gain of 1.07 dB, and transmission within 1.2 dB of the Shanon limit.

References

[1] G.D.Forney and L.F. Wei “Multidimensional constellations—Part I: Introduction, figures of Merit, and generalized cross constellations”.

IEEE Journal on selected areas in communications, Vol. 7, No. 6, pp. 877–892, August 1989.

[2] R.G. Gallager "Information Theory and Reliable Communication". *Wiley, New York* 1968.

[3] A.R. Calderbank and L.H. Ozarow "Nonequiprobable signaling on the Gaussian channel" *IEEE Transaction on information theory*, Vol. 36, No. 4, pp. 726–740, July 1990.

[4] F.R. Kschischang and S.Pasupathy "Optimal nonuniform signaling for Gaussian channels". *IEEE Transactions on Information Theory*, vol. 39, No.3, May 1993, pp. 913-929.

[5] P.Robertson, T.Woerz "Bandwidth efficient turbo trellis-coded modulation using punctured component codes". *Journal on Selected Areas in Communications*, vol. 16, No.2, Feb. 1998, pp. 206-218.

[6] U.Wachsmann, R.Fischer and J.B.Huber "Multilevel codes: theoretical concepts and practical design rules ". *IEEE Transactions on information Theory*, vol. 45, No.5, May 1999, pp. 1361-1391.

[7] S. Le Goff , A. Glavieux and C. Berrou "Turbo codes and high efficiency modulation," *in Proc. of IEEE ICC' 94,* New Orleans, LA, May 1994, pp. 645-649

[8] C. Fragouli and D. Wessel "Turbo encoder design for symbol-interleaved parallel concatenated trellis-coded modulation". *IEEE Transactions on Communications*, vol. 49, No.3, March 2001, pp. 425-430.

[9] S. Benedetto and G. Montorsi "Versatile bandwidth-efficient parallel and serial turbo-trellis-coded modulation". *IEEE International symposium on Turbo Codes 2000*, pp. 201-208.

[10] S. Benedetto and G. Montorsi "Generalized concatenated codes with interleavers ". *IEEE International symposium on Turbo Codes 1997*, pp. 32-39.

[11] F.Mo, S.C.Kwatra and J.Kim "Analysis of puncturing pattern for high rate turbo codes ". *MILCOM 1999. IEEE Military Communications. Conference Proceedings vol. 1*, pp. 547-550.

[12] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate", *IEEE Trans. Inf. Theory*, p. 284–287, Mar. 1974.

[13] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes", *Proc. 1993 Int. Conf. Comm.*, p. 1064–1070