Performance Analysis of a Multilevel Coded Modulation System

Yosef Kofman, Ephraim Zehavi, *Senior Member, IEEE*, and Shlomo Shamai (Shitz), *Senior Member, IEEE*

Abstract— **A modified version of the multilevel coded modulation scheme of Imai & Hirakawa is presented and analyzed. In the transmitter, the outputs of the component codes are bit interleaved prior to mapping into 8-PSK channel signals. A multistage receiver is considered, in which the output amplitudes of the Gaussian channel are soft limited before entering the second and third stage decoders. Upper bounds and Gaussian approximations for the bit error probability of every component code, which take into account errors in previously decoded stages, are presented. Aided by a comprehensive computer simulation, it is demonstrated in a specific example that the addition of the interleaver and soft limiter in the third stage improves its performance by 1.1 dB at a bit error probability of 10^{-5}, and that the multilevel scheme improves on an Ungerboeck's code with the same decoding complexity. The rate selection of the component codes is also considered and a simple selection rule, based on information theoretic arguments, is provided.**

I. INTRODUCTION

Ungerboeck, in his pioneering work [1], showed that by combining channel coding and modulation into one entity by means of a trellis code, it is possible to achieve a remarkable coding gain as compared to an uncoded modulation system with the same spectral efficiency and data rate. The core of Ungerboeck's scheme, which is, in fact, fundamental in almost every coded modulation system, is the mapping by set partitioning in which the redundant signal set is partitioned into smaller subsets with increasing intra-set Euclidean distances. Independently, Imai & Hirakawa [2] proposed a coded modulation scheme based on a multilevel binary code. In this method the channel signal set is successively binary partitioned, using the set partitioning rule, where the binary labels of the edges from one level of the partition chain to the next are encoded by independent binary codes. The multilevel scheme enables the usage of a suboptimal multistage decoder as was originally proposed in [2] (see also [3-9]), which demonstrates performance/complexity advantages over the maximum likelihood decoder, since its overall complexity is proportional only to the decoding complexity sum of the component codes and not their product (see [3-9]).

The performance of a multilevel scheme with a multistage decoder has been approximated mainly by the Asymptotic Coding Gain (ACG) as compared to an appropriate uncoded modulation system, or by the effective coding gain which accounts also the error coefficients of the code [3-8]. When block codes have been used as component codes, certain upper bounds on the error event probability have been also evaluated [10, 11]. However, as was pointed out in [7], it is not simple to extend these error event bounds to provide upper bounds on the decoded bit error probability because of the interaction between decoders in various stages. In fact, because of the side information passed among subsequent stages, the coding channel of each stage is neither memoryless nor Gaussian. The main focus of the present paper is to present a rigorous analysis of the decoded bit error probability for a multilevel system which incorporates convolutional codes at each stage and a multistage decoder at the receiver end.

The multilevel scheme, presented in the next section, is a modified version of [2]. The modification is expressed in two aspects: 1) Addition of interleaver/deinterleaver pairs in the second and third stages, which prevents the appearance of error bursts in the decoded sequences of the component convolutional codes. These interleavers also increase the diversity level of the system, a characteristic which appears to be advantageous in fading channels [12-14]. 2) Soft limiting the output amplitudes of the channel before entering the second and third stage decoders. These soft limiters resemble practical receivers with restricted dynamic range, and also diminish the effect of incorrect side information passed among subsequent stages, a fact which turns to be advantageous in decoding the third stage code. In this paper we specialize to binary convolutional codes and an 8-PSK signal set, but the generalization to other codes or signal sets is straight forward.

The error performance analysis of the proposed multistage decoder is presented in section III. Upper bounds on the bit error probability which take into account the side information passed among subsequent stages, are derived for every component code. In addition, approximations for these bounds which use the assumption that the coding channel of each stage is Gaussian, ignoring, thus, the statistical dependence between the noise samples and the sequences of the previously decoded stages, are presented. Recently, an error performance analysis of a multistage decoder was reported in [5, 15]. It

Paper approved by Stephen G. Wilson, the Editor for Coding Theory and Applications of the IEEE Communications Society. Manuscript received: January 17, 1991; revised March 26, 1992. This work was supported by the Ministry of Communications, Chief Scientist, grant no. 050703, and it includes portions of a D.Sc. dissertation by Y. kofman, submitted to the Senate of the Technion in June 1992. This paper was presented in part at the 1990 Bilkent International Conference on New Trends in Communication, Control and Signal Processing, Ankara, Turkey, July 1990.

Y. Kofman was with the Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel. He is now with Teledata Communication Ltd., P.O.B. 2003, Herzlia 46120, Israel.

E. Zehavi is with Qualcomm, Inc., 10555 Sorrento Valley Road, San Diego, CA 92121.

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel.

IEEE Log Number 9400912.

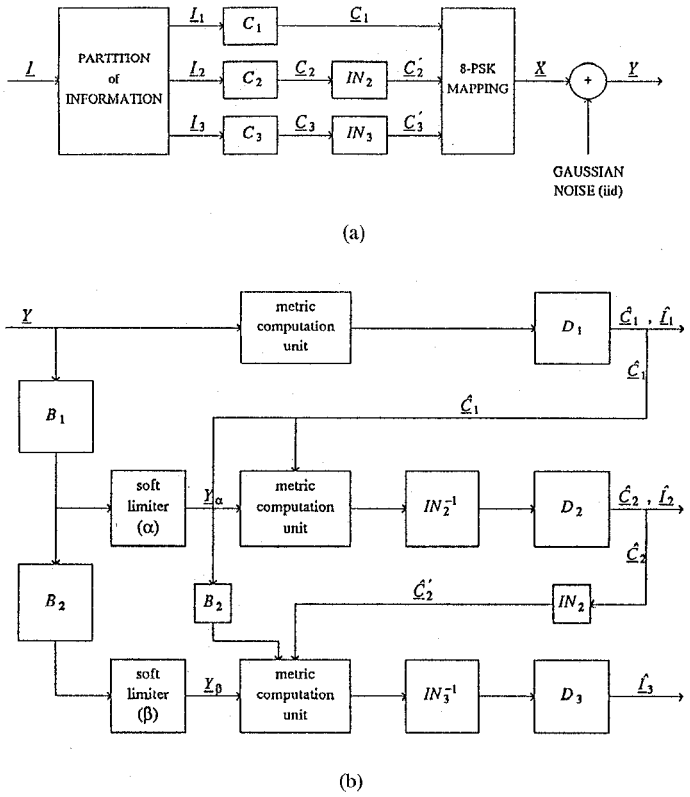

Fig. 1. (a) Transmitter block diagram. (b) Receiver block diagram.

seems, however, that the aforementioned Gaussian assumption was used as the basis of the analysis. In section IV, an example along with results of a comprehensive computer simulation of different multistage decoders are presented. It is demonstrated that the addition of the interleaver and soft limiter in the third stage improves its performance by 1.1 dB at a bit error probability of 10^{-5}, and that the error performance of the proposed scheme improves on Ungerboeck's codes with the same decoding complexity. The issue of choosing the rates and constraint lengths of the component codes is addressed in section V, where a simple rule for selecting the rates, which is based on information theoretic arguments for the aggregate rate of multiuser systems, is provided.

II. SYSTEM DESCRIPTION AND PRELIMINARIES

The transmitter, described in Fig. 1.a, is based on the scheme in [2]. A binary information sequence, $\underline{I}$, is partitioned into three binary subsequences $\underline{I}_1$, $\underline{I}_2$ and $\underline{I}_3$, where each subsequence is convolutionally encoded by an independent binary component code, denoted by $\boldsymbol{C}_i$, $i = 1,2,3$. The rate of code C_i is $R_i = k_i/n_i$, and its output code sequence is denoted by $\underline{C}_i$. The code sequences $\underline{C}_2$ and $\underline{C}_3$ are bit interleaved prior to mapping into channel signals. The associated bit interleavers IN_2 and IN_3 are assumed to be independent and ideal, that is their output sequences $\underline{C}'_2$ and $\underline{C}'_3$, respectively, are statistically independent of each other and consist of independent and identically distributed (iid) bits.

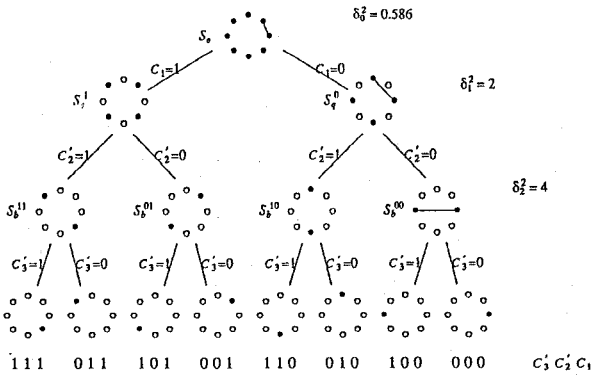

Fig. 2. Set partitioning of an 8-PSK signal set.

Three bits, two from the bit interleavers' output and another from the code sequence $\underline{C}_1$, are mapped synchronously into one of 8-PSK channel signals. The 8-PSK signal set is denoted by S_o and given by $S_o \triangleq \{a_k = \sqrt{E_s}\exp(j2\pi k/8),\ k = 0,1,\ldots,7\}$. The mapping, illustrated in Fig. 2, follows the set partitioning rule of [1] and [2]. The binary labels of the edges from one level of the partition chain to the next, denoted by C_1, C'_2 and C'_3, are components of the sequences $\underline{C}_1$, $\underline{C}'_2$ and $\underline{C}'_3$, respectively. The subscripts o, q and b of the subsets stand for octal, quaternary and binary, respectively, and refer to the number of elements in each subset, whereas the superscripts define the rightmost coded bits conveyed by every channel signal in the subset. The squared Euclidean subset distance of S_o is $\delta_o^2 = 0.586\,E_s$. C_1 determines a QPSK signal set, for which the intra-set squared Euclidean distance is $\delta_q^2 = 2\,E_s$. C_1 and C'_2 specify a BPSK signal set, for which the intra-set squared Euclidean distance is $\delta_b^2 = 4\,E_s$.

The channel considered in this paper is the discrete-time, memoryless Gaussian channel, where $\underline{X}$ and $\underline{Y}$ stand, respectively, for the input and the corresponding output sequences. Due to the independent convolutional codes and bit interleavers, the sequence $\underline{X}$ consists of independent and identically distributed channel signals. The noise samples are independent Gaussian complex random variables with zero mean and variance $N_0/2$ in each dimension. The overall rate, designated by R, is given by $R = R_1 + R_2 + R_3$ *bits/channel signal*. The signal to noise ratio (SNR) is E_s/N_0 and E_b/N_0, standing for the received energy per bit to noise ratio, equals E_s/N_0R.

The multistage decoder, described in Fig. 1.b, is a modified version of the one proposed in [2]. Each stage consists of a metric computation unit and a Viterbi decoder, denoted by D_i, $i = 1,2,3$, which produces the estimated information sequence $\hat{\underline{I}}_i$ and code sequence $\hat{\underline{C}}_i$. In addition, symbol deinterleavers, denoted by IN_2^{-1} and IN_3^{-1}, are introduced in stage 2 and stage 3, respectively. These symbol deinterleavers perform the inverse operation of the corresponding bit interleavers in the transmitter. The estimated code sequence, $\hat{\underline{C}}_2$, is passed again through the ideal bit interleaver, IN_2, in order to

preserve synchronization with the appropriate received signal. The overall operation of the interleaver/deinterleaver pairs eliminate error bursts in the side information sequences passed among subsequent stages. Since Viterbi decoder is sensitive to error bursts, the addition of the interleaver/deinterleaver pairs improves the error performance of the system at most E_b/N_0 values. The delay buffers, B_1 and B_2, store the received sequence, $\underline{Y}$, until decoding in previous stages is completed. Soft amplitude limiters with thresholds α and β, respectively, are introduced in the second and third stages, where $\underline{Y}_\alpha$ and $\underline{Y}_\beta$ stand, respectively, for their output sequences. These soft limiters are devices with a unity gain up to a threshold, whence the output amplitude is clipped and the phase is left unchanged. They resemble practical systems with a restricted dynamic range and mitigate, as well, the effect of incorrect side information in the second and third decoding stages.

In each stage, the side information determines the signal set upon which the metrics are calculated according to the set partitioning structure (illustrated in Fig. 2). For the predetermined signal set, the suboptimal metric associated with the code bit '0' or '1' is minus of the corresponding squared Euclidean distance. For a rate k/n component convolutional code, the sum of n consecutive metrics is calculated for every branch metric in the trellis diagram of the code.

More specifically, the metric, evaluated in stage 1, is denoted by $m_1(Y\ ;\ C_1)$ and defined by

$$m_1(Y\ ;\ C_1) = -\min_{X \in S_q^{C_1}} |Y - X|^2 \quad , \quad C_1 = 0, 1 \tag{1}$$

where Y and X are the received and possible transmitted channel signals, respectively, and $C_1 = 0, 1$ is the code bit associated with the subset $S_q^{C_1}$ (see Fig. 2). The metric evaluated in stage 2 and associated with the side information bit $\hat{C}_1$ is denoted by $m_2(Y_\alpha\ ;\ C_2'|\hat{C}_1)$ and defined by

$$m_2(Y_\alpha\ ;\ C_2'|\hat{C}_1) = -\min_{X \in S_b^{C'_2\ \hat{C}_1}} |Y_\alpha - X|^2 \quad , \quad \hat{C}_1, C_2' = 0, 1 \tag{2}$$

where Y_α is the soft limiter output for which the amplitude is limited to be below a threshold α, and $S_b^{C'_2\ \hat{C}_1}$ $(C_2', \hat{C}_1 = 0, 1)$ is the BPSK subset, determined by C_2' and $\hat{C}_1$ (see Fig. 2). Similarly, the metric $m_3(Y_\beta\ ;\ C'_3|\hat{C}_2', \hat{C}_1)$ associated with the side information bits $\hat{C}_2'$ and $\hat{C}_1$ is given by

$$m_3(Y_\beta\ ;\ C'_3|\hat{C}_2', \hat{C}_1) = -|Y_\beta - X|^2 \tag{3}$$

where X is the transmitted channel signal determined by C'_3, $\hat{C}_2'$ and $\hat{C}_1$ (see Fig. 2), and Y_β is the soft limiter output for which the amplitude is limited to be below a threshold β.

Note that for the case in which the side information is incorrect, the metrics are calculated on a signal set which is rotated by 45° (for stage 2) and 45° or 90° or 135° (for stage 3) with respect to the true one and, therefore, are wrong. The soft limiters introduced in stages 2 and 3 diminish this effect by bounding the maximal possible value that each wrong metric may take.

We conclude this section by considering the decoding complexity per information bit of the system, which is the usual basis for comparing different codes and coding schemes. Usually the decoding complexity is related only to the number of states of the convolutional code. However, the computation of the branch metric may be a heavy task for low rate codes which are expected to be used in the first stage. Therefore, for the sake of a fair comparison, the decoding complexity is defined by the number of binary comparisons and binary metric additions required for decoding one bit of information. A straight forward calculation, which can be found in [9] and [12], reveals that the decoding complexity per information bit of a rate k/n convolutional code with 2^ν states is give by $(2^\nu \cdot (2^k - 1) + 2^n - 2)/k$. For implementation simplicity, high rate codes are often selected to be punctured convolutional codes, based on rate 1/2 convolutional codes [16]. For these codes, the decoding complexity is almost the same as the decoding complexity of the original code and is given by 2^ν. The overall decoding complexity of the system, denoted by L, is given by the weighted sum $L = \sum_{i=1}^{3}(R_i\ L_i)/R$ where R is the overall rate of the system and L_i is the decoding complexity of each stage.

A trellis coded modulation scheme is a natural candidate for a comparison with the scheme in this paper. The decoding complexity per information bit of a trellis code, based on a rate $(n-1)/n$ convolutional code with 2^ν states, is given by $(2^\nu\ (2^{n-1} - 1))/(n - 1)$.

III. ERROR PERFORMANCE

The performance of a multilevel coded modulation system, which employs a multistage decoder, has been commonly approximated by the asymptotic coding gain as compared to an uncoded system with the same data rate and spectral efficiency [3, 4, 6, 7, 8], or by the effective coding gain, which incorporates the error coefficients of the components codes [6 - 8]. For coded 8-PSK specifically, denote by d_{H1}, d_{H2} and d_{H3} the free Hamming distance of component codes $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ and $\boldsymbol{C}_3$, respectively. Then, the ACG, expressed in dB, as compared to an uncoded QPSK modulation system with the same energy is given by

$$ACG = 10 \cdot log(\min\ \{\ 0.293\, d_{H1}\ ,\ d_{H2}\ ,\ 2\, d_{H3}\ \})\ . \tag{4}$$

In this section we present rigorous upper bounds along with approximations for the average bit error probability of each component code, which account for errors in the side information passed among subsequent stages. Denote by P_{b1}, P_{b2} and P_{b3} the average bit error probability of codes $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ and $\boldsymbol{C}_3$, respectively. Then, the average bit error probability of the overall system, denoted by P_b, is given by

$$P_b = \frac{P_{b1} \cdot R_1 + P_{b2} \cdot R_2 + P_{b3} \cdot R_3}{R}\ . \tag{5}$$

The coding channel, namely the channel from encoder to decoder of each stage, is memoryless due to the independent

and ideal interleaver/deinterleaver pairs introduced in the transmitter and the receiver. This implies that the multivariate probability density functions $f(\underline{Y}|\underline{X})$, $f(\underline{Y}_\alpha|\underline{X},\underline{\hat{C}}_1)$ and $f(\underline{Y}_\beta|\underline{X},\underline{\hat{C}}_2',\underline{\hat{C}}_1)$ associated with the first, second and third stages, respectively, can be factored into a product of scalar conditional probability density functions. Explicitly stated, for stage 1, $f(\underline{Y}|\underline{X}) = \prod_j f(Y_j|X_j)$, for stage 2, $f(\underline{Y}_\alpha|\underline{X},\underline{\hat{C}}_1) = \prod_j f(Y_{\alpha j}|X_j,\hat{C}_{1j})$, and for stage 3, $f(\underline{Y}_\beta|\underline{X},\underline{\hat{C}}_2',\underline{\hat{C}}_1) = \prod_j f(Y_{\beta j}|X_j,\hat{C}_{2j}',\hat{C}_{1j})$.

Considering a general decoder, the estimated code sequence at its output reflects the statistical properties of the noise in the channel. In our case, since a stage decoder is employed, the estimated code sequences of previously decoded stages are passed as side information to subsequent stages. Consequently, these side information sequences, which carry some statistical knowledge about the noise in the channel, alter the statistical nature of the noise of the *coding channel* of each stage. For the second stage, for example, the event that the side information is correct, that is $C_1 = \hat{C}_1$, implies that the noise in the channel was "small" at the vicinity of $\hat{C}_1$, whereas the event that the side information is wrong, namely $C_1 \neq \hat{C}_1$, implies that the noise was "large". As a result, the noise seen by the decoder in the second stage can not be treated for granted as being Gaussian. Taking advantage of the soft limiters, however, enables us to bypass this obstacle and to obtain rigorous upper bounds for the decoded average bit error probability in the second and third stages.

A first order approximation to the coding channels of the second and third stages, which is also presented in the paper, is obtained by ignoring the statistical knowledge carried by the side information and considering the noise of the coding channels as Gaussian, after all. The approximation, based on the above assumption, is referred to, throughout, as the Gaussian approximation. It seems that the Gaussian assumption was also the basis for the error performance analysis in [5, 15], where in [5], a so called "maximum likelihood" metric was employed, as well, at the multistage decoder.

In order to simplify notations, we consider only the coding channels in the following derivation. This means that the interleaver/deinterleaver pairs are disregarded and the side information sequences are taken as $\underline{\hat{C}}_1$ and $\underline{\hat{C}}_2$ (instead of $\underline{\hat{C}}_2'$), while considering them as being comprised of iid bits.

A. Upper Bounds

1) Stage 1: The derivation of the upper bound on the average bit error probability of the first stage follows the approach used in [7] for QAM signal sets. A similar approach can be found also in [9] and in [12] for a fading channel. Consider first the pairwise error probability $P(\underline{C}_1 \to \underline{\hat{C}}_1)$ of choosing the code sequence $\underline{\hat{C}}_1$ instead of the actually transmitted code sequence $\underline{C}_1$. Suppose that $\underline{C}_1$ is conveyed by the specific channel signal sequence $\underline{X}$. Due to the squared Euclidean distance metric of (1), used in the first stage, the decision between $\underline{C}_1$ and $\underline{\hat{C}}_1$ can be viewed as a conventional maximum likelihood decoding over a Gaussian channel, where $\underline{X}$ is the transmitted sequence and the possible decoded sequences are those sequences which may convey $\underline{\hat{C}}_1$. The decision region of each sequence is defined, in the usual manner, as the set of points that are closest, in terms of the squared Euclidean distance, to that sequence. Clearly, the union of the decision regions of the channel signal sequences that are closest to $\underline{X}$ covers the entire space of channel signal sequences, except the decision region of $\underline{X}$. Each 8-PSK channel signal has two neighbors equally close to it, that is, those signals that are 45 degrees apart. Therefore, if the Hamming distance between $\underline{C}_1$ and $\underline{\hat{C}}_1$ is d, there are 2^d possible channel signal sequences closest to $\underline{X}$ with squared Euclidean distance $0.586\,E_s\,d$. Hence, applying the union bound, $P(\underline{C}_1 \to \underline{\hat{C}}_1)$ is upper bounded by

$$P(\underline{C}_1 \to \underline{\hat{C}}_1) \le 2^d Q\left(\sqrt{\frac{0.586 \cdot E_s\, d}{2N_0}}\right) \tag{6}$$

where $Q(\cdot)$ is the Gaussian integral function defined by $Q(\delta) = 1/\sqrt{2\pi}\int_\delta^\infty \exp(-t^2/2)dt$.

Note that the above derivation of the error probability of the first stage may be applied to any M-PSK signal set by just replacing the argument $0.586\,E_s$ in the square root of (6) by the appropriate squared Euclidean distance between nearest neighbor channel signals.

Due to (6), each error coefficient of the code $\boldsymbol{C}_1$ is multiplied by the factor 2^d, thus, increasing the effective error coefficients of the code. The appearance of this factor is explained by the fact that the code sequences $\underline{C}_2$ and $\underline{C}_3$ are seen by the decoder in the first stage as uncoded sequences since it receives no help in decoding from the second and third stages (see [7-9]). For low rate codes with large free Hamming distance, which are expected to be used in the first stage, the factor 2^d may be significant and loosen the bound even at medium E_b/N_0.

The bound in (6) depends only on the Hamming distance between the sequences $\underline{C}_1$ and $\underline{\hat{C}}_1$ and not on their specific structure. Therefore, it may be assumed, without loss of generality, that the all-zero code sequence is transmitted, and it is possible to incorporate the conventional generating function approach [17]. Thus, using the relation $Q(\sqrt{\delta+\gamma}) \le Q(\sqrt{\delta})\exp(-\gamma/2)$ (see [17]) for bounding $P(\underline{C}_1 \to \underline{\hat{C}}_1)$ in (6), the average bit error probability P_{b1} is upper bounded for a rate $R_1 = k_1/n_1$ convolutional code $\boldsymbol{C}_1$ by

$$P_{b1} \le \frac{1}{k_1} Q\left(\sqrt{\frac{0.586 \cdot E_s d_{H1}}{2N_0}}\right) \cdot \exp\left(\frac{0.586 \cdot E_s d_{H1}}{4N_0}\right) \cdot \left.\frac{\partial T_1(D,I)}{\partial I}\right|_{I=1, D=2\exp\left(-\frac{0.586E_s}{4N_0}\right)} \tag{7}$$

where d_{H1} is the free Hamming distance of code $\boldsymbol{C}_1$ and $T_1(\cdot,\cdot)$ is its generating function.

In order to calculate the error performance of the second and third stage decoders, the reliability of the side information has to be also evaluated. The average error probability of a reencoded symbol (as opposed to a decoded bit) belonging to the component code C_1 is denoted by P_{c1} and is upper bounded, as shown in Appendix B, by

$$P_{c1} \le \frac{1}{n_1} Q\left(\sqrt{\frac{0.586 \cdot E_s d_{H1}}{2N_0}}\right) \cdot \exp\left(\frac{0.586 \cdot E_s d_{H1}}{4N_0}\right) \cdot D \frac{\partial T_1(D,I)}{\partial D}\Big|_{I=1, D=2\exp\left(-\frac{0.586E_s}{4N_0}\right)} \quad (8)$$

2) Stage 2: The pairwise error probability $P(\underline{C}_2 \to \underline{\hat{C}}_2)$ of choosing the code sequence $\underline{\hat{C}}_2$ instead of the transmitted code sequence $\underline{C}_2$ is given by

$$P(\underline{C}_2 \to \underline{\hat{C}}_2) = \sum_{\underline{C}_1, \underline{\hat{C}}_1} P(\underline{C}_1) P(\underline{\hat{C}}_1 | \underline{C}_1) \cdot P\left(\sum_j m_2(Y_{\alpha j}; \hat{C}_{2j} | \hat{C}_{1j}) - m_2(Y_{\alpha j}; C_{2j} | \hat{C}_{1j}) \ge 0 | \underline{C}_1, \underline{\hat{C}}_1, \underline{C}_2\right) \quad (9)$$

where $\underline{C}_1$ is the transmitted code sequence of component code C_1, $\underline{\hat{C}}_1$ is the corresponding estimated code sequence passed as side information, $Y_{\alpha j}$ is the soft limiter output signal and $m_2(\cdot\,;\cdot\,|\,\cdot)$ is the metric defined in (2). Applying the Chernoff bounding technique, $P(\underline{C}_2 \to \underline{\hat{C}}_2)$ is upper bounded by

$$P(\underline{C}_2 \to \underline{\hat{C}}_2) \le \sum_{\underline{C}_1, \underline{\hat{C}}_1} P(\underline{C}_1) P(\underline{\hat{C}}_1 | \underline{C}_1) \cdot E\left(E\left(\exp\left\{\lambda \sum_j \left[m_2(Y_{\alpha j}; \hat{C}_{2j} | \hat{C}_{1j}) - m_2(Y_{\alpha j}; C_{2j} | \hat{C}_{1j})\right]\right\} | \underline{X}, \underline{\hat{C}}_1\right) | \underline{C}_1, \underline{\hat{C}}_1, \underline{C}_2\right) \quad (10)$$

where "E" is the statistical expectation operation, $\underline{X}$ is a possible channel signal sequence conveying the code sequences $\underline{C}_1$ and $\underline{C}_2$, and $\lambda \ge 0$ is the Chernoff bound parameter to be determined. Due to the symmetrical signal set and the fact that the codes are linear, it is assumed without loss of generality (see [9]) that the transmitted code sequences are $\underline{C}_1 = \underline{C}_2 = \underline{0}$ and the associated channel signal sequence is $\underline{X} = \underline{a}_0$. In addition, since $\underline{X}$ depends on $\underline{\hat{C}}_1$ through $\underline{C}_1$, the outer expectation is independent of $\underline{\hat{C}}_1$. Using the fact that the coding channel of stage 2 is memoryless due to the interleaver/deinterleaver pair introduced at this stage, and dropping the time dependence j in the notation, the upper bound in (10) takes the form

$$P(\underline{C}_2 \to \underline{\hat{C}}_2) \le \left[\sum_{\hat{C}_1=0,1} P(\hat{C}_1 | C_1 = 0) \cdot E\left(\exp\left\{\lambda\left[m_2(Y_\alpha; 1|\hat{C}_1) - m_2(Y_\alpha; 0|\hat{C}_1)\right]\right\} | a_0, \hat{C}_1\right)\right]^d, \quad (11)$$

where d is the Hamming distance between $\underline{C}_2$ and $\underline{\hat{C}}_2$.

There are two cases to consider in evaluating the expectation in (11). The first is the one in which the side information is wrong, i.e., $\hat{C}_1 = 1$ and $C_1 = 0$, and the second is the one in which the side information is correct, i.e., $\hat{C}_1 = C_1 = 0$. Each such event carries some statistical knowledge about the corresponding noise sample in the channel, and therefore the memoryless coding channel seen by the Viterbi decoder in stage 2 (and also in stage 3) is not necessarily Gaussian any more. Consider first the expectation for the case in which the side information is wrong. Since this expectation involves the *unknown* conditional probability density function $f(Y_\alpha | a_0, \hat{C}_1 = 1)$, we upper bound the metric difference by the maximal value it takes on. From definition (2) of the metric used in the second stage, and by using simple geometrical arguments, it is easily seen that the maximal value of the metric difference is $2\alpha\sqrt{E_s}$, where α is the soft limiter threshold amplitude. Taking $\lambda = 1/2N_0$ (see Appendix A), the expectation in the above case is upper bounded by

$$E\left(\exp\left\{\lambda\left[m_2(Y_\alpha; 1|1) - m_2(Y_\alpha; 0|1)\right]\right\} | a_0, \hat{C}_1 = 1\right) \le \exp\left(\frac{\alpha\sqrt{E_s}}{N_0}\right). \quad (12)$$

Consider next the case in which the side information is correct. In order to get rid of the conditioning on $\hat{C}_1$ in (11), we employ the upper bound

$$f(Y_\alpha | a_0, \hat{C}_1 = 0) \le \frac{f(Y_\alpha | a_0)}{P(\hat{C}_1 = 0 | C_1 = 0)}$$

on the conditional probability density function, where $f(Y_\alpha | a_0)$ is easily derived, as shown in Appendix A, from the conditional probability density function $f(Y|a_0)$ of the received signal before the soft limiter. Using the fact that the arguments of the expectation in (11) are all positive and taking $\lambda = 1/2N_0$ (see Appendix A), the expectation in (11) is bounded for the case of correct side information by

$$E\left(\exp\left\{\lambda\left[m_2(Y_\alpha; 1|0) - m_2(Y_\alpha; 0|0)\right]\right\} | a_0, \hat{C}_1 = 0\right) \le 2\exp\left(-\frac{E_s}{2N_0}\right) \cdot \left[1 - QM\left(\sqrt{E_s/N_0}, \frac{\alpha}{\sqrt{N_0/2}}\right)\right] + 2J_2 \quad (13)$$

where $QM(\cdot,\cdot)$ is the Marcum Q-function defined by $Q(\delta_1, \delta_2) = \int_{\delta_2}^{\infty} t \exp\left(-(t^2 + \delta_1^2)/2\right) I_0(\delta_1 t) dt$, α is the soft limiter threshold amplitude, and J_2 is the integral given by

$$J_2 = \int_{\alpha}^{\infty} \frac{2r}{N_0} \exp\left(-\frac{r^2+E_s}{N_0}\right) I_0\left(\sqrt{\frac{E_s}{N_0}\,\frac{4r^2-4r\alpha+2\alpha^2}{N_0}}\right) dr \quad , \tag{14}$$

with $I_0(\cdot)$ standing for the modified Bessel function of order zero. Inserting (12) and (13) into (11), and using the notation $P(\hat{C}_1 = 1|C_1 = 0) = P_{c1}$ where P_{c1}, bounded in (8), is the average error probability of a reencoded symbol belonging to code $\boldsymbol{C}_1$, the following upper bound results:

$$P(\underline{C}_2 \to \hat{\underline{C}}_2) \le \left(2\exp(-\frac{E_s}{2N_0}) \cdot \left[1 - QM\left(\sqrt{E_s/N_0}, \frac{\alpha}{\sqrt{N_0/2}}\right)\right] + 2J_2 + P_{c1}\exp\left(\frac{\alpha\sqrt{E_s}}{N_0}\right)\right)^d \triangleq D_2^d \, . \tag{15}$$

Finally, applying the generating function approach [17], the average bit error probability P_{b2} is upper bounded for a rate $R_2 = k_2 / n_2$ convolutional code $\boldsymbol{C}_2$ by

$$P_{b2} \le \frac{1}{k_2} \cdot \frac{\partial T_2(D,I)}{\partial I}\bigg|_{I=1, D=D_2} \tag{16}$$

where $T_2(\cdot,\cdot)$ is the generating function of code C_2 and D_2 is defined in (15). The average error probability of a reencoded symbol, denoted by P_{c2}, is upper bounded as shown in Appendix B by

$$P_{c2} \le \frac{1}{n_2} \cdot D\,\frac{\partial T_2(D,I)}{\partial D}\bigg|_{I=1, D=D_2} \, . \tag{17}$$

3) Stage 3: The error probability for the third stage is obtained following steps that are similar to those used in the derivation of the error probability for the second stage. Therefore, only one point of difference is emphasized, where a detailed derivation is given in [9]. The pairwise error probability $P(\underline{C}_3 \to \hat{\underline{C}}_3)$ is given by an expression similar to (9). The side information passed to stage 3 consists of the sequences $\hat{\underline{C}}_1$ and $\hat{\underline{C}}_2$. Therefore, the probability $P(\hat{C}_1, \hat{C}_2|C_1, C_2)$ has to be evaluated. Since the coding channels of the various stages are not independent from each other, due to the statistical knowledge conveyed by the side information, the above probability can not be factored into the product $P(\hat{C}_1|C_1)\,P(\hat{C}_2|C_2)$. Instead, the upper bound $P(\hat{C}_1, \hat{C}_2|C_1, C_2) \le 2\min\{P(\hat{C}_1|C_1), P(\hat{C}_2|C_2)\}$ is employed [9]. Consequently, for the case of wrong side information, the exponent is multiplied by $2(P_{c1} + P_{c2} + \min\{P_{c1} + P_{c2}\})$. The pairwise error probability $P(\underline{C}_3 \to \hat{\underline{C}}_3)$ is, thus, bounded by

$$P(\underline{C}_3 \to \hat{\underline{C}}_3) \le \left[\exp\left(-\frac{E_s}{N_0}\right) \cdot \left[1 - \exp\left(-\frac{\beta^2}{N_0}\right)\right] + J_3 + 2\cdot\left(P_{c1} + P_{c2} + \min\left\{P_{c1}, P_{c2}\right\}\right) \cdot \exp\left(\frac{2\beta\sqrt{E_s}}{N_0}\right)\right]^d \triangleq D_3^d \tag{18}$$

where β is the soft limiter threshold amplitude, P_{c1} and P_{c2} are the average error probabilities of reencoded symbols belonging to codes $\mathbf{C}_1$ and $\mathbf{C}_2$, respectively, and J_3 is the integral given by

$$J_3 = \int_{\beta}^{\infty} \frac{2r}{N_0} \exp\left(-\frac{r^2+E_s}{N_0}\right) I_0\left(\frac{2\sqrt{E_s}}{N_0}(r-\beta)\right) dr \quad . \tag{19}$$

The upper bound on P_{b3} is obtained simply by replacing k_2 and D_2 with k_3 and D_3, respectively, in (16).

B. Gaussian Approximations

In order to simplify the presentation of the Gaussian approximations, we consider a receiver without soft limiters in stages 2 and 3. Thus, the Viterbi decoders in these stages use conventional soft decisions.

The pairwise error probability $P(\underline{C}_2 \to \hat{\underline{C}}_2)$ is shown in Appendix C to be approximated by

$$P(\underline{C}_2 \to \hat{\underline{C}}_2) \le 2^d \sum_{m=0}^{d} \binom{d}{m} (P_{c1})^m (1-P_{c1})^{d-m} \cdot Q\left(\sqrt{\frac{(d-m)^2\,E_s}{dN_0}}\right) \tag{20}$$

where d is the Hamming distance between $\underline{C}_2$ and $\hat{\underline{C}}_2$. The derivation of the Gaussian approximation for the third stage is similar to that of the second stage, therefore, the details are provided in [9], where only the concluding result concerning the pairwise error probability $P(\underline{C}_3 \to \hat{\underline{C}}_3)$ is given here as follows:

$$P(\underline{C}_3 \to \hat{\underline{C}}_3) \approx \sum_{\substack{m_1,m_2,m_3,m_4 \\ m_1+m_2+m_3+m_4=d}} \frac{d!}{m_1!\,m_2!\,m_3!\,m_4!} \cdot (P_{c1})^{m_2+m_4}(1-P_{c1})^{m_1+m_3}(P_{c2})^{m_3+m_4}(1-P_{c2})^{m_1+m_2} P(r \ge 0) \tag{21}$$

where $0 \le m_1, m_2, m_3, m_4 \le d$, d is the Hamming distance between $\underline{C}_3$ and $\hat{\underline{C}}_3$, and r is a Gaussian random variable with mean $-2E_s[2m_1 + \sqrt{2}(m_2 - m_4)]$ and variance $8\,d\,E_s N_0$.

The Gaussian approximation on the average bit error probability is obtained in the usual manner [17] as

$$P_b \le \frac{1}{k} \sum_{i=1}^{\infty} \sum_{d=d_H}^{\infty} i\,a(d,i)\,P_d \tag{22}$$

where d_H is the free Hamming distance of either code $\boldsymbol{C}_2$ or code $\boldsymbol{C}_3$, $a(d,i)$ is the number of paths in the trellis diagram of

the convolutional code which diverge from the all-zero path for Hamming distance d and carry i 1's in their unmerged segment, and P_d is the associated pairwise error probability given in either (20) or (21). The average probability of a reencoded symbol, P_{c2}, is approximated by the expression (B.1) in Appendix B.

It is possible to obtain approximations which make use of a compact form of the generating function of the code [17] by upper bounding the pairwise error probabilities in (20) and (21) with the usage of the Chernoff bounding technique, or to upper bound the Q-function in (20) by an exponential term. This subject, however, is omitted here for the sake of brevity.

C. Discussion

We conclude this section by examining the upper bounds and approximations obtained for the various stages. The multistage decoder is sequential in nature and, therefore, the influence of an untight bound at some stage propagates through the subsequent stages and may result in bounds which are even less tight. Since the code $\boldsymbol{C}_1$ is expected to be powerful, i.e., a code with a large free Hamming distance, its effective error coefficients would be significantly large (see the discussion following (6)). Consequently, the above propagation would begin, in many cases of interest, already at the first stage.

The above problematic nature of the bounds holds in general. When the bounds in the second and third stages are considered, other questions are raised. The bounding technique in these stages relies upon bounding the metric difference by the maximal value it takes on when the side information is wrong. Although this bounding technique seems to be too pessimistic at first glance, the degradation of the bounds due to it would be small in many cases of interest, since the probability of a wrong side information bit is usually small.

There are, however, two cases for which the degradation is expected to be significant; the cases for which the codes $\boldsymbol{C}_2$ or $\boldsymbol{C}_3$ have free Hamming distance of either 1 (uncoded case) or 2. The reasons for a possible significant degradation in the above cases is demonstrated for the second stage. Consider the expression (9) of the pairwise error probability in the second stage. For the uncoded case and when the side information bit is wrong, upper bounding the metric difference by its maximal value, which is clearly greater than zero, results in the trivial bound of 1 for the pairwise error probability. For the second case in which the free Hamming distance of code $\boldsymbol{C}_2$ is 2, consider the event for which one bit of side information is wrong and the other is correct. Clearly, the sum of the metric differences associated with this event is at least zero, and the corresponding pairwise error probability is bounded by 1, as well. Consequently, the overall pairwise error probability in these cases would be proportional to the error probability P_{c1} of the side information in the first stage, which, on its own right, does not necessarily exhibit the correct behavior of the bound in the second stage. Note that when the Chernoff bounding technique is used for bounding the pairwise error probability, the above situation clearly becomes worse.

For the third stage, the degradation in the aforementioned cases would be even larger since there are more side information sequences which lead to the trivial bound, and due to the factor of 2 of P_{c1} and P_{c2} in (18). Note that for the third stage, free Hamming distance of 2 already results in an ACG of 6.02 dB, which is sufficient for all practical cases.

Consider now the Gaussian assumption. It seems that the this assumption is a good *working* assumption since there is only a slight degradation, if any, as compared to the receiver incorporating soft decisions, as is seen by simulation results for a specific example, presented in section IV. This is expected as the statistical dependence between a certain decoded decision and its aligned noise sample is small for powerful codes. Still, since a degradation is observed, at least for the second stage, the Gaussian approximations can not be considered as strict upper bounds, even if they appear to be so in various figures when compared to computer simulation results.

IV. AN EXAMPLE AND SIMULATION RESULTS

In this section we examine the bounds and approximations obtained in the previous section through an example accompanied by comprehensive computer simulations. The convolutional codes chosen as component codes are listed in Table I. For each code, we have indicated its rate (R_i), number of states, Hamming distance (d_{Hi}), decoding complexity (L_i), Asymptotic Coding Gain (ACG), and the suitable reference. The codes $\boldsymbol{C}_2$ and $\boldsymbol{C}_3$ are punctured convolutional codes, based on rate 1/2 convolutional codes. The overall rate of the system is $R = R_1 + R_2 + R_3 = 2$ *bits / channel signal*, the ACG is 6.02 dB according to (4), and the overall decoding complexity per information bit of the multistage decoder is $L = 20.92$.

Before presenting the bounds and approximations, we describe the various computer simulations performed with the selected codes. We performed computer simulations for the system proposed in this paper and a system which represents the Gaussian assumption. For comparison purposes, we also simulated a multilevel system without soft limiters, a system with no interleavers, and a system in which the side information is perfect. In all the simulations performed, the data sequence was, with no loss of generality, the all-zero sequence, and the Viterbi decoders operated with practically infinite quantization and path length.

The interleaver in the second stage was implemented as a block interleaver with 288 rows and 288 columns. This block interleaver has a depth of 14.4 constraint lengths with respect to code $\boldsymbol{C}_1$, and a span of 9.6 constraint lengths with respect to code $\boldsymbol{C}_2$. Thus, the interleaver can be considered for practical purposes as ideal. In the third stage, the interleaver was implemented as a pseudo random interleaver, where the sequence length interleaved each time was $288^2 = 82944$ bits long. It is not clear that this interleaver is long enough to be considered ideal, but by comparing the simulation results using this interleaver with those of the Gaussian assumption it is seen that the degradation, if any, is small.

The system which represents the Gaussian assumption was implemented by simulating each stage independently of the others in order to eliminate the statistical dependence between the side information and the noise in the coding channel. Each bit in the iid side information sequences used in the second

TABLE I
Component Codes' Specifications

Code	Rate	States	Hamming Distance	Complexity	ACG (dB)	Reference
C_1	1/4	32	18	46	7.22	[18]
C_2	5/6	32	4	32	6.02	[16]
C_3	11/12	4	2	4	6.02	[19]

TABLE II
Soft Limiters Threshold Parameters

E_b/N_0 (dB)	6.25	6.5	6.75	7.0	7.25	7.5	7.75	8.0	8.25	8.5	8.75	9.0
$\alpha/\sqrt{E_s}$	0.66	0.66	0.72	0.75	0.8	0.85	0.9	0.93	0.95	1.05	1.05	1.05
$\beta/\sqrt{E_s}$	0.24	0.24	0.27	0.3	0.32	0.34	0.36	0.37	0.38	0.39	0.4	0.41

and third stages was obtained by flipping a coin with a probability of success which was equal to the error probability of reencoded symbols simulated in the previously decoded stages of the multistage decoder, where the multistage decoder incorporated soft decisions in every stage.

We present now the bounds and Gaussian approximations along with the relevant simulation results. The upper bounds on the average bit error probability of each component code and the overall proposed system are shown in Fig. 3. The thresholds parameters, α and β, listed in Table II were taken so as to optimize the bounds for every E_b/N_0. The bound on P_{b1} was calculated by using 21 coefficients from the generating function (see [18]), and the bound on P_{b2} was calculated by using 8 coefficients (see [16]). For P_{b3}, the *full* generating function was evaluated. Also shown in Fig. 3 are the corresponding simulation results. The threshold parameters taken were $\alpha = 0.66\sqrt{E_s}$ and $\beta = 0.24\sqrt{E_s}$. No attempt has been made to optimize these parameters to get the best possible simulation results. It is observed from the graphs that the bound on P_{b1} is not tight enough even for a bit error probability of 10^{-5}, and it becomes very loose for error probabilities larger than that. The above phenomenon is due to the large effective error coefficients of code C_1, attributed to the factor 2^d in (6) (see section III). In particular, since the free Hamming distance of the convolutional code used in the first stage is 18, the first error coefficient of code C_1 is multiplied by 2^{18} which is a significant factor even for medium and large E_b/N_0 values. Note that the bound for the first stage is based on Q-functions and the bounding is mainly due to the union bound, which is the basis for most known bounds evaluated for convolutional codes incorporating soft decisions. Note also that since the decoder is a sequential one, the fact that the bound for the error probability of the side information, P_{c1}, is not tight enough causes a drastic threshold effect in the bounds for P_{b2} and P_{b3} already at an error probability of $5 \cdot 10^{-5}$.

In order to examine the tightness of the bound on P_{b2}, an

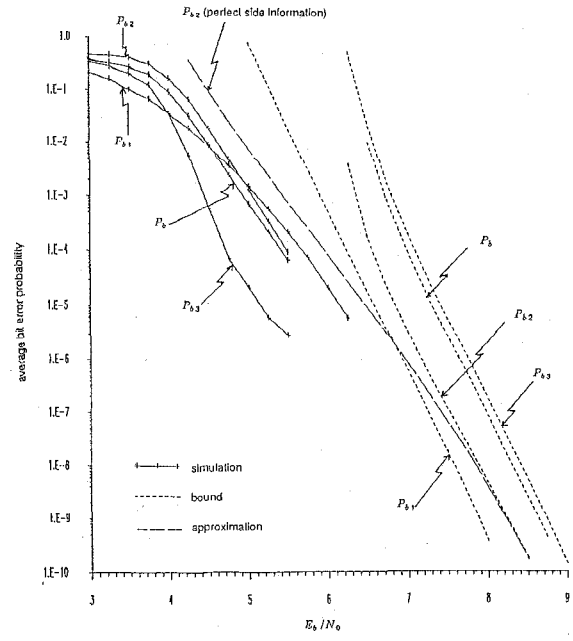

Fig. 3. Upper bounds (P_{b1}, P_{b2}, P_{b3}, P_b) and simulation results (P_{b1}, P_{b2}, P_{b3}, P_b) for the multilevel system with interleavers and soft limiters, and an approximation for P_{b2}.

approximation, which considers only correct side information sequences and uses soft decisions, is also presented in Fig. 3. The Chernoff bound-based average bit error probability for this approximation is obtained simply by taking $P_{c1} = 0$ and $\alpha = \infty$ in (16). A comparison between the graphs of the exact bound and the above approximation reveals that they are in a good agreement especially for medium and large values of E_b/N_0. Since for low values of E_b/N_0 the side information becomes significant, it is concluded that the untightness of the bound is mainly due to the Chernoff bounding technique and the fact that the bound on P_{c1} is not adequately tight. The bound for P_{b3} is very loose due to the fact that code C_3 has a free Hamming distance of 2 (see the end of section III for a discussion), and the fact that the bounds on the side information from previously decoded stages, P_{c1} and P_{c2}, are not tight. As was pointed out in section III, this causes the bound to be worse than either the bound for P_{c1} or the bound for P_{c2}. Taking a code C_3 with a free Hamming distance greater than 2 would result in a much better bound, but as can be seen by the simulation results, the error performance of the third stage is already significantly superior to the other two. Note that although the ACG of code C_1 is 7.22 dB whereas the ACG of codes C_2 and C_3 is only 6.02 dB, their error performance at medium and large E_b/N_0 is better than the error performance of code C_1. This implies the important conclusion that the codes should not be selected on an equal ACG basis, but, instead, the first code should be the most powerful one since it is the less protected in terms of the intra-set squared Euclidean distance.

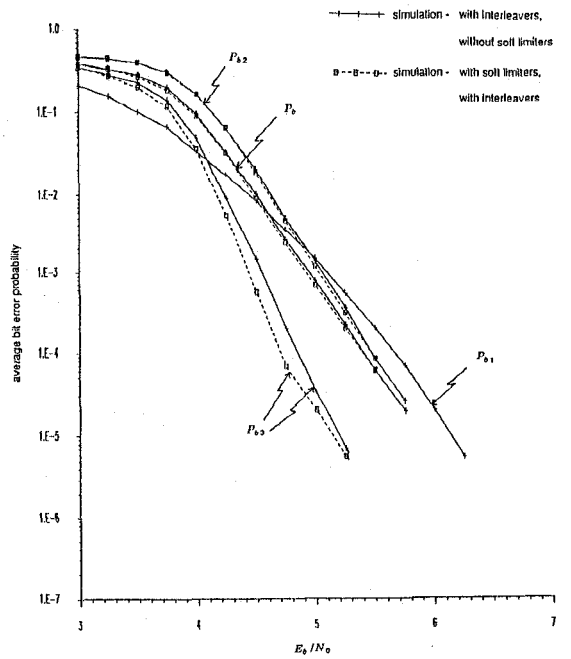

Fig. 4. Simulation results (P_{b1}, P_{b2}, P_{b3}, P_b) for the system with/without soft limiters and with interleavers.

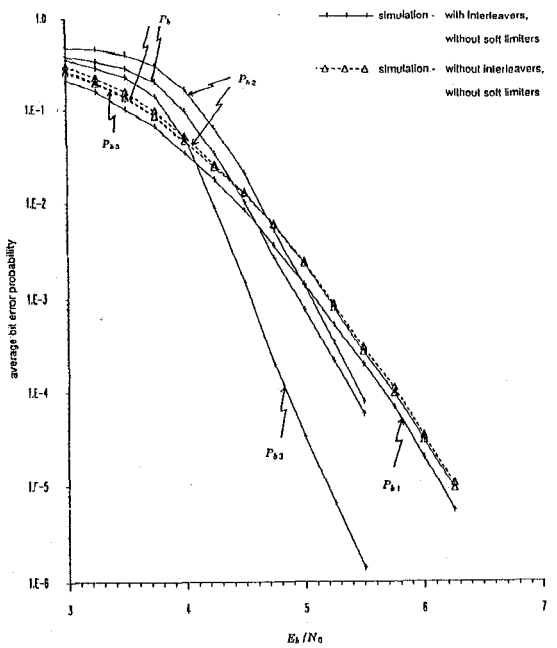

Fig. 5. Simulation results (P_{b1}, P_{b2}, P_{b3}, P_b) for the system with/without interleavers and without soft limiters.

In order to examine the effect of the soft limiters we simulated also a coding system without the soft limiters. These simulation results along with those of the system with the soft limiters are depicted in Fig. 4. It is observed that the improvement due to the addition of the soft limiters is negligible in stage 2, whereas for stage 3, there is an improvement for low

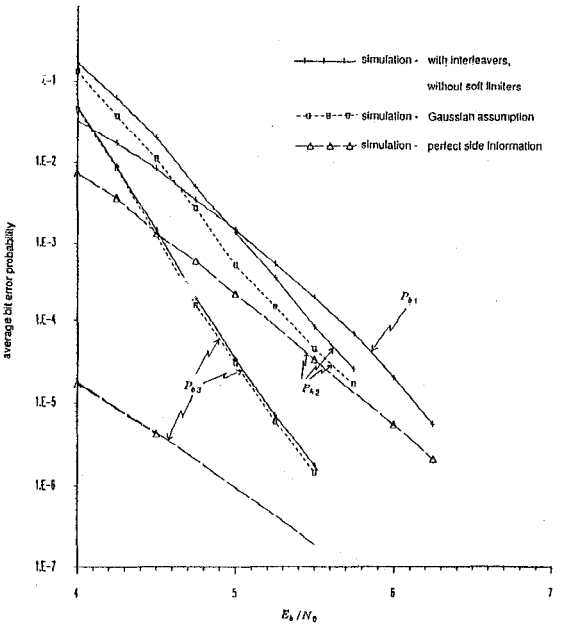

Fig. 6. Simulation results for the system without soft limiters and with interleavers (P_{b1}, P_{b2}, P_{b3}), for the Gaussian assumption (P_{b2}, P_{b3}), and for the system with perfect side information (P_{b2}, P_{b3}).

E_b/N_0 of up to 0.15 dB (for $E_b/N_0 = 4.75$ *dB*). The above phenomenon is explained by observing the set partitioning structure (see Fig. 2) and noting that the metrics calculated in the second stage are less effected by an incorrect side information than those metrics calculated in the third stage.

Another comparison, shown in Fig. 5, is performed between the coded system with the interleavers and a system with no interleavers, where both systems have no soft limiters. It is observed that the addition of the interleaver in stage 2 reduces the error performance for E_b/N_0 less than 5 dB and is advantageous only for E_b/N_0 greater than 5 dB where the improvement itself is modest. On the other hand, the addition of the interleaver in the third stage improves the error performance significantly for almost any E_b/N_0. For example, an improvement of 1.05 dB is exemplified at error probability of 10^{-5}. The performance of the system without the interleavers is governed by the first stage, since the occurrence of an error burst in this stage necessarily inflicts an error in the second and the third stages. It is worthwhile mentioning that a simulation performed for a system without interleavers but with soft limiters yields almost no difference as compared to the simulation results of the system without interleavers and soft limiters. This observation is also explained by the assertion that the occurrence of an error burst implies an error in a subsequent stage almost independently of the metrics being used.

In the next two figures we examine the Gaussian assumption, its validity as a *working* assertion, and the resulted Gaussian approximations. In Fig. 6 we present simulation results of a system which resembles the Gaussian assumption along with simulation results of the system with no soft limiters. It is demonstrated that a difference in performance exists only in the second stage at low E_b/N_0 and, in any case, this difference is no more than about 0.25 dB. The fact that the difference in

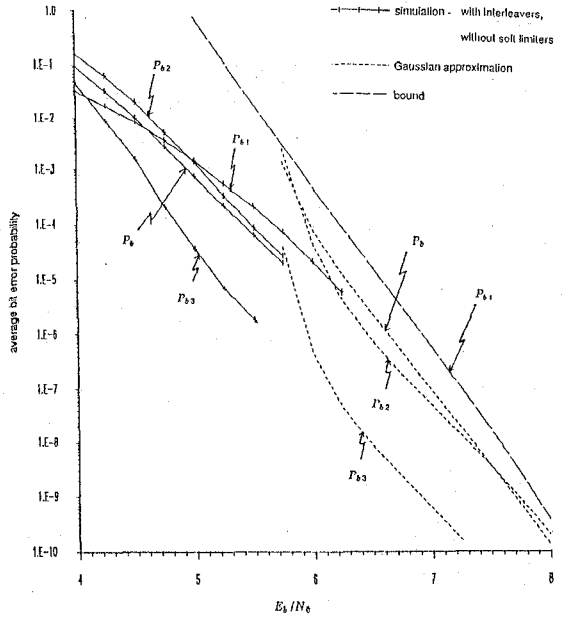

Fig. 7. Simulation results (P_{b1}, P_{b2}, P_{b3}, P_b) for the system with interleavers and without soft limiters, an upper bound for P_{b1}, and Gaussian approximations (P_{b2}, P_{b3}, P_b).

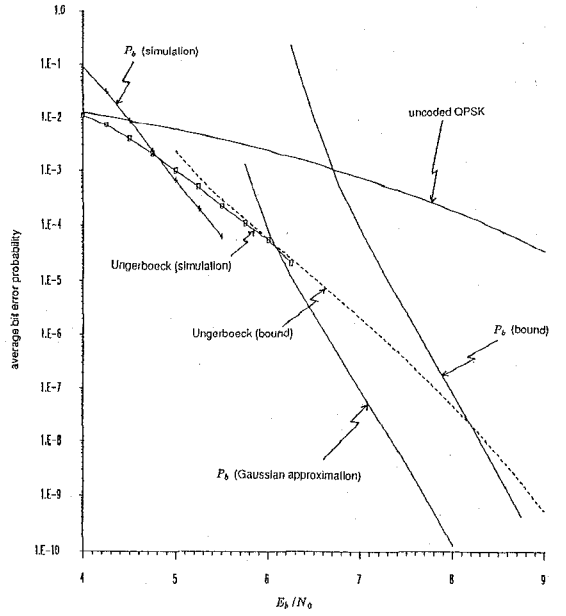

Fig. 8. An overall average bit error probability for the multilevel system with interleavers and soft limiters (an upper bound and simulation), for the Gaussian assumption (an approximation), for a 16 states Ungerboeck code (an upper bound and simulation), and for an uncoded QPSK.

the second stage is in favor of the Gaussian assumption is due to the elimination of the statistical dependence between the side information and the aligned noise samples in the system which represents the Gaussian assumption. On the other hand, the lack of difference in the third stage possibly occur since the error performance in this stage is dominated by the wrong metrics which results from a wrong side information (see Fig. 2 and section III), where the effect of a change in the statistics of the noise samples is rather small. Note, however, that since a degradation is still observed, the Gaussian approximations which results from the Gaussian assumption can not be considered as strict upper bounds, even if they appear to be so in the next figure. In fact, the reason that the approximations appear as bounds is mainly due the underlying union bound on which they are based.

Another important issue is the conditions under which the side information sequences can essentially be considered perfect. To examine this, a system which incorporates only correct side information sequences has been simulated and the results are also depicted in Fig. 6. For the third stage, in order to assess the performance at low error probabilities, an upper bound is also shown. For stage 2, it is observed that the side information may eventually be considered perfect at a bit error probability of 10^{-5}, whereas for stage 3, there is still 1 dB difference at that error probability. Moreover, for stage 3 the side information can not be considered perfect even for a bit error probability of 10^{-6}. Note that for the system without interleavers the difference for stage 3 is 2.05 dB at an error probability of 10^{-5}. The results obtained so far put in question the validity of the ACG or even the effective coding gain, as discussed at the beginning of section III, as relevant parameters for approximating the performance of multilevel codes, since they do not take into account the side information passed among subsequent stages.

After showing that the Gaussian assumption is indeed a reasonable working assertion, we present in Fig. 7 the Gaussian approximations for the bit error probability of the second and third stages and the overall system (without soft limiters). The approximation for P_{b3} was obtained here by taking only 4 coefficients in (22) (instead of the full generating function). For comparison purposes, the associated simulation results of the system with interleavers, but without soft limiters, are also shown. It is observed that the Gaussian approximations provide a good estimate of the error probability for medium and large E_b/N_0, and that for low E_b/N_0 they suffer from the same problems as the rigorous bounds for P_{b2} and P_{b3}. It is interesting to observe that the largest difference between the bound on P_{b2} (see Fig. 3) and its associated Gaussian approximation is at most 0.7 dB. At a bit error probability of 10^{-9} it can be assumed that the side information, passed to the second stage, is perfect, as is seen from Fig. 6. Therefore, by comparing the bound on P_{b2} and the approximation at this error probability, it is concluded that about 0.5 dB of the difference is due to the Chernoff bounding technique used for bounding P_{b2}. For the third stage, the Gaussian approximation, based on the Chernoff bounding technique, was also evaluated, and a difference of about 0.8 dB was observed from the approximation given in (21). Thus, the bounding technique used to bound P_{b3} (eq. (18)) contributes another degradation of about 0.8-0.9 dB.

We conclude this section by comparing the results obtained so far with an appropriate Ungerboeck code with 16 states, which has decoding complexity per information bit of 24 and ACG of 4.13 dB [1]. An upper bound on the bit error probability of this code, taken from [20], is shown in Fig. 8 along with

simulation results. Also shown are the upper bound and the simulation results for the proposed system, the Gaussian approximation and the bit error probability of an uncoded QPSK. By comparing the graphs, it is observed that for roughly the same decoding complexity the multilevel scheme considered in the example improves on the 16 states Ungerboeck code for E_b/N_0 larger than 4.75 dB. Note, however, that due to the multistage decoder and the added interleaver/deinterleaver pairs, the decoding delay and memory storage required by the multilevel scheme are considerably larger.

V. RATES SELECTION VIA INFORMATION THEORETIC ARGUMENTS

One of the main design parameters of a multilevel coded modulation system, which employs a multistage decoder, is the selection of the component binary codes. We propose a simple rule for selecting the component codes' rates, which is based on information theoretic arguments for the aggregate rate of multiuser systems. Once the rates are determined, the problem of choosing the number of memory elements of each component convolutional code can be resolved using the upper bounds or approximations on the error probability of the various component codes, given in section III.

The multilevel scheme modeled as a multiuser system is described in Fig. 9. C_i $(i = 1,2,3)$ is the binary input of channel i, where the inputs are iid. Using a simple identity of average mutual information functionals [21], along with the fact that the channel signal X is uniquely mapped by C_1, C_2 and C_3, yield

$$\begin{aligned} I(Y;X) &= I(Y;C_1,C_2,C_3) \\ &= I(Y;C_1) + I(Y;C_2|C_1) + I(Y;C_3|C_2,C_1) \end{aligned} \tag{23}$$

where $I(\cdot;\cdot)$ is the average mutual information functional and $I(\cdot;\cdot|\cdot)$ and $I(\cdot;\cdot|\cdot,\cdot)$ are the conditional ones, all expressed in bits per channel signal. When using an ideal multistage decoder which employs, of course, a maximum likelihood metric in each stage, $I(Y;C_1)$, $I(Y;C_2|C_1)$, and $I(Y;C_3|C_2,C_1)$ are the maximal theoretical component rates possible in each stage [22]. When a maximum likelihood decoder is used instead of the multistage decoder, the maximal theoretical component rate for stage 1, stage 2, and stage 3 is given by $I(Y;C_1|C_2,C_3)$, $I(Y;C_2|C_1,C_3)$, and $I(Y;C_3|C_2,C_1)$, respectively [22]. The expressions for the component rates of the multistage decoder reflect its sequential nature, whereas those for the maximum likelihood decoder exhibit the fact that decoding is performed simultaneously for all the stages, and, therefore, each stage "receives help" in decoding from the other stages. Clearly, the maximal component rates possible with the maximum likelihood decoder are larger than those possible with the multistage decoder. Nevertheless, since $I(Y,X)$ is still the maximal theoretical *aggregate* rate possible with a maximum likelihood decoder [22], expression (23) establishes the optimality of an ideal multistage decoder, as far as the rate issue is concerned. It is worthwhile mentioning that $I(Y,X)$, being the average mutual information between the input and the output of the channel, is also the maximal theoretical rate of Ungerboeck-like schemes [1].

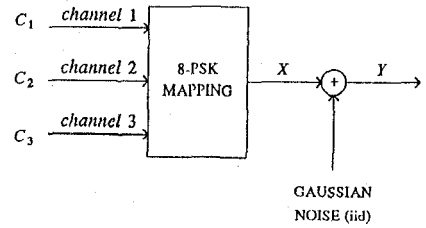

Fig. 9. Multiuser system model of a multilevel code.

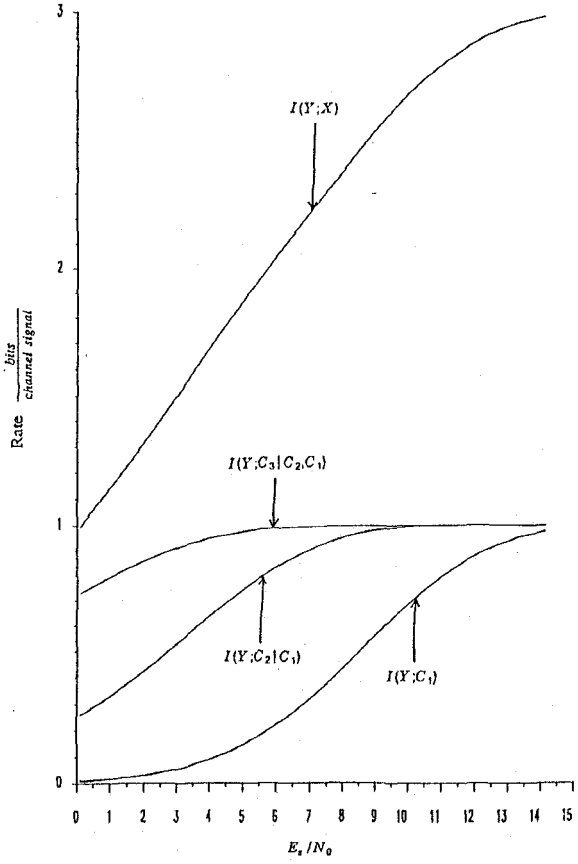

Fig. 10. Maximal component codes' rates versus E_s/N_0 for a multilevel 8-PSK coded modulation system with a multistage decoder.

Based on the above discussion, the suggested rule for selecting the component codes' rates goes as follows: For an overall rate, R, evaluate E_s/N_0 for which $I(Y;X) = R$. For this E_s/N_0, determine the component codes' rates to be $R_1 = I(Y;C_1)$, $R_2 = I(Y;C_2|C_1)$ and $R_3 = I(Y;C_3|C_2,C_1)$. The relations $I(Y;C_1) \le I(Y;C_2|C_1) \le I(Y;C_3|C_2,C_1)$ are in agreement with the observation that the further we proceed in the partition chain the less powerful component codes are needed. Note that the discussion which preceded the suggested rule was quite general and independent of a specific channel or signal set. An application of the suggested rule in a Rayleigh fading channel, for example, can be found in [12]. It is worthwhile mentioning that another approach for selecting the memory of each component code, which is based on random coding bounds, has been proposed in [23]. For the

8-PSK signal set, $I(Y, C_1)$, $I(Y, C_2|C_1)$, $I(Y, C_3|C_2, C_1)$, and $I(Y, X)$ versus E_s/N_0 are depicted in Fig. 10.

In conclusion, we examine the specific example presented in the previous section. According to the proposed rule, for an overall rate $R = 2$, the component codes' rates are calculated to be $R_1 = 0.2$, $R_2 = 0.813$, and $R_3 = 0.987$, which are reasonably close to the actual rates selected in the example.

VI. CONCLUSIONS

A multilevel 8-PSK coded modulation scheme, which employs bit interleavers and soft limiters in various stages of the transmitter and the multistage decoder, has been presented. Upper bounds and Gaussian approximations on the bit error probability, which take into account the side information passed among subsequent stages, have been derived for every component code. For a specific example, it has been demonstrated that the addition of the interleavers and soft limiters improves the performance of the third stage by 1.1 dB. It has been shown also that the proposed multilevel scheme improves on Ungerboeck code with the same decoding complexity at the expense of increasing decoding delay. The issue of the component codes' rates has been also addressed, and a simple rule for selecting these rates, based on information theoretic arguments, has been proposed.

An addition of an interleaver/deinterleaver pair in the first stage will enable the usage of the coding system proposed in this paper in a bursty noise environment. The direct proportion between Hamming distances of the component codes and the squared Euclidean distance of the overall code, as expressed in (4), along with the inherent diversity provided by the interleaving approach, were found to be advantageous in a Rayleigh fading channel, as was reported in [12].

APPENDIX A
DERIVATION OF (13) and (14)

In order to perform the expectation in (13), the conditional probability density function of the soft limiter output signal, Y_α, of the second stage has to be evaluated. Given that the transmitted channel signal is a_0, the conditional probability density function of the received signal, Y, is given in polar coordinates by (see [24, pp. 167])

$$f(r, \phi|a_0) = \frac{r}{\pi N_0} \exp\left(-\frac{r^2 + E_s - 2\sqrt{E_s} r \cos\phi}{N_0}\right)$$

where $r \geq 0$ and $0 \leq \phi \leq 2\pi$ are the amplitude and phase of Y, respectively. Since the soft limiter only limits the amplitude of Y and does not alter its phase, the conditional probability density function of Y_α is given in polar coordinates by

$$f(r_\alpha, \phi|a_0) = \begin{cases} f(r, \phi|a_0) & , \quad 0 \leq r_\alpha < \alpha \\ \int_\alpha^\infty f(r, \phi|a_0)\, dr & , \quad r_\alpha = \alpha \\ 0 & , \quad r_\alpha > \alpha \end{cases} \tag{A.1}$$

where $0 \leq r_\alpha \leq \alpha$ and $0 \leq \phi \leq 2\pi$ are the amplitude and phase of Y_α, respectively. Next, bounding the metric difference (see [9]) and using the inequality $e^{\max(a,b)} \leq e^a + e^b$, the expectation in (13) is bounded by

$$E\left(\exp\{\lambda[m_2(Y_\alpha; 110) - m_2(Y_\alpha; 010)]\}|a_0\right) \tag{A.2}$$

$$\leq E\Big(\exp\{2\lambda\sqrt{E_s} r_\alpha \max\{\cos(\phi - \frac{\pi}{2}) - \cos(\phi),$$

$$\cos(\phi - \frac{3\pi}{2}) - \cos(\phi)\}\}|a_0\Big)$$

$$\leq 2E\left(\exp\{2\lambda\sqrt{E_s} r_\alpha [\sin(\phi) - \cos(\phi)]\}|a_0\right)$$

where the factor 2 in the second inequality is due to the phase symmetry of the expectation. Making use of the probability density function given in (A.1), and after some straight forward calculations, the expectation (A.2) is then expressed in terms of the following integrals

$$E\left(\exp\{2\lambda\sqrt{E_s} r_\alpha [\sin(\phi) - \cos(\phi)]\}|a_0\right) \tag{A.3}$$

$$= \int_0^\alpha \frac{2r}{N_0} \exp\left(-\frac{r^2 + E_s}{N_0}\right) I_0\left(2\sqrt{E_s} r \sqrt{\lambda^2 + (\frac{1}{N_0} - \lambda)^2}\right) dr$$

$$+ \int_\alpha^\infty \frac{2r}{N_0} \exp\left(-\frac{r^2 + E_s}{N_0}\right) I_0\left(2\sqrt{E_s}\sqrt{\lambda^2\alpha^2 + (\frac{r}{N_0} - \lambda\alpha)^2}\right) dr.$$

Denoting $A^2 = 4E_s\left(\lambda^2 + (\frac{1}{N_0} - \lambda)^2\right)$, the first integral in (A.3) can be further simplified to yield

$$\int_0^\alpha \frac{2r}{N_0} \exp\left(-\frac{r^2 + E_s}{N_0}\right) I_0(rA)\, dr \tag{A.4}$$

$$= 2\exp\left(-\frac{E_s}{N_0} + \frac{A^2 N_0}{4}\right) \cdot$$

$$\left[QM(A\sqrt{N_0/2}, 0) - QM(A\sqrt{N_0/2}, \frac{\alpha}{\sqrt{N_0/2}})\right]$$

$$= 2\exp(-2\lambda E_s(1 - \lambda N_0)) \cdot \left[1 - QM\left(A\sqrt{N_0/2}, \frac{\alpha}{\sqrt{N_0/2}}\right)\right]$$

where $QM(\cdot,\cdot)$ is the Marcum Q-function and the identity $QM(\delta, 0) = 1$ has been used in the second equality.

Since the Chernoff bound is valid for any $\lambda \geq 0$, we choose $\lambda = 1/2N_0$, which optimizes the expression $2\exp\left(-2\lambda E_s(1 - \lambda N_0)\right)$, that would have resulted as an upper bound replacing (A.4) if the side information has always been correct and soft decision ($\alpha = \infty$) has been incorporated. This choice implies that $A = \sqrt{2E_s}/N_0$, which upon

substitution in (A.4) yields

$$E\left(\exp\{\lambda[m_2(Y_\alpha;110)-m_2(Y_\alpha;010)]\}|a_0\right) \qquad \text{(A.5)}$$

$$\le 2\exp\left(-\frac{E_s}{2N_0}\right)\cdot\left[1-QM\left(\sqrt{E_s/N_0},\frac{\alpha}{\sqrt{N_0/2}}\right)\right]$$

$$+2\int_{\alpha}^{\infty}\frac{2r}{N_0}\exp\left(-\frac{r^2+E_s}{N_0}\right)I_0\left(\sqrt{\frac{E_s}{N_0}\frac{(4r^2-4r\alpha+2\alpha^2)}{N_0}}\right)dr.$$

The desired derivation is concluded by letting the integral in (A.5) be designated by J_2.

APPENDIX B
A BOUND ON THE AVERAGE ERROR PROBABILITY OF A REENCODED SYMBOL

In this Appendix, an upper bound on the average error probability of a reencoded symbol, denoted by P_c, is derived for convolutional codes. Let us denote by $T(D)$ the transfer function of a rate k/n convolutional code, given by $T(D)=\sum_{d=d_H}^{\infty}a(d)D^d$ where d_H is the free Hamming distance of the code, and $a(d)$ is the number of paths in the trellis diagram of the convolutional code which diverge from the all-zeros path for Hamming distance d. In a similar way to the derivation of the bit error probability [17], P_c is upper bounded by the expected number of errors in the reencoded symbols caused by an incorrect path diverging at any node in the trellis diagram, divided by the total number of reencoded symbols generated in that node. Since an incorrect path with Hamming distance d carries d reencoded symbols, and since n reencoded symbols are generated at any node, P_c is upper bounded by

$$P_c \le \frac{1}{n}\sum_{d=d_H}^{\infty} d\,a(d)\,P_d \qquad \text{(B.1)}$$

where P_d is the associated pairwise error probability.

Suppose that P_d may be upper bounded by $P_d \le Z^d$, then using the transfer function $T(D)$, P_c is bounded by

$$P_c \le \frac{1}{n}D\frac{\partial T(D)}{\partial D}\Big|_{D=Z}. \qquad \text{(B.2)}$$

For stage 1, the bound (8) on P_{c1} is achieved by employing the bound [17] $Q(\sqrt{\delta+\gamma}) \le Q(\sqrt{\delta})\exp(-\frac{\gamma}{2})$ in (6), and taking $Z=2\exp(-0.586E_s/4N_0)$. For stage 2, the bound (17) on P_{c2} is reached by taking

$$Z=2\exp\left(-\frac{E_s}{2N_0}\right)\cdot\left[1-Q\left(\sqrt{E_s/N_0},\frac{\alpha}{\sqrt{N_0/2}}\right)\right]+2J_2$$

$$+P_{c1}\exp\left(\frac{\alpha\sqrt{E_s}}{N_0}\right).$$

APPENDIX C
DERIVATION OF THE GAUSSIAN APPROXIMATION FOR THE SECOND STAGE

The pairwise error probability $P(\underline{C}_2 \to \hat{\underline{C}}_2)$ is given in (9). It is assumed, without loss of generality, that $\underline{C}_1=\underline{C}_2=\underline{0}$ and $\underline{X}=\underline{a}_0$. We denote by d the Hamming distance between $\underline{C}_2$ and $\hat{\underline{C}}_2$. Since for $C_{2j}=\hat{C}_{2j}$ the metric difference is zero, we assume without loss of generality that all sequences considered in this appendix are of length d, where the sequence $\hat{\underline{C}}_2$ is the all-one sequence. An upper bound on $P(\underline{C}_2 \to \hat{\underline{C}}_2)$ is obtained by upper bounding the metric difference. For $\hat{C}_{1j}=0$, the metric difference is bounded for every Y_j by (see [9])

$$m_2(Y_j,110)-m_2(Y_j,010) \le \max\{2Re[Y_j^*(a_2-a_0)], 2Re[Y_j^*(a_6-a_0)]\} \qquad \text{(C.1)}$$

and for $\hat{C}_{1j}=1$ the metric difference is bounded by

$$m_2(Y_j,111)-m_2(Y_j,011) \le \max\{2Re[Y_j^*(a_3-a_1)], 2Re[Y_j^*(a_7-a_1)]\}. \qquad \text{(C.2)}$$

For a side information sequence of length d, we assume that there are m $(0\le m\le d)$ places for which the side information is wrong, namely $\hat{C}_{1j}=1$, and $(d-m)$ places for which it is correct, namely $\hat{C}_{1j}=0$. The pairwise error probability $P(\underline{C}_2 \to \hat{\underline{C}}_2)$ is bounded, then, by taking a union bound over all $\binom{d}{m}\cdot 2^d$ possible permutations for selecting $(d-m)$ elements out of the terms $2Re(Y^*(a_2-a_0))$ and $2Re(Y^*(a_6-a_0))$, and m elements out of the terms $2Re(Y^*(a_3-a_1))$ and $2Re(Y^*(a_7-a_1))$. Let us examine one such permutation for which out of m wrong side information bits, we select the bound $2Re(Y^*(a_3-a_1))$ m_1 $(0\le m_1\le m)$ times and $(m-m_1)$ times the bound $2Re(Y^*(a_7-a_1))$. For this permutation, using the Gaussian assumption and given that the transmitted channel signal is a_0, the bound on the sum of the metric differences, denoted by u, is a Gaussian random variable with variance $\sigma_u^2=2dN_0|a_2-a_0|^2$ and mean $\mu_u=-(d-m)|a_2-a_0|^2-m_1(|a_3-a_0|^2-|a_1-a_0|^2)$. Now, using the fact that $Q(\cdot)$ is an increasing function we obtain

$$P(u\ge 0)=Q\left(\frac{(d-m)|a_2-a_0|^2+m_1(|a_3-a_0|^2-|a_1-a_0|^2)}{\sqrt{2dN_0|a_2-a_0|^2}}\right)$$

$$\le Q\left(\sqrt{\frac{(d+m_1-m)^2|a_2-a_0|^2}{2dN_0}}\right) \qquad \text{(C.3)}$$

$$\le Q\left(\sqrt{\frac{(d-m)^2|a_2-a_0|^2}{2dN_0}}\right)$$

where the first inequality is due to the relation

$|a_3 - a_0|^2 - |a_1 - a_0|^2 \geq |a_2 - a_0|^2$, and the second inequality results since $0 \leq m_1 \leq m$ and $0 \leq m \leq d$.

The bound in (C.3) is independent of m_1 and is valid for any permutation which is comprised of $(d-m)$ correct side information bits and m wrong ones. Since there is an over all number of $\binom{d}{m} \cdot 2^d$ such permutations, then by union bounding over all permutations and taking $|a_2 - a_0|^2 = 2E_s$, $P(\underline{C}_2 \rightarrow \hat{\underline{C}}_2)$ is approximated by the expression given in (20).

REFERENCES

[1] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. Information Theory,* Vol. IT-28, No. 1, pp. 55-67, Jan. 1982.

[2] H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error-Correcting Codes," *IEEE Trans. Information Theory,* Vol. IT-23, No. 3, pp. 371-377, May 1977.

[3] S.I. Sayegh, "A Class of Optimum Block Codes in Signal Space," *IEEE Trans. Communications,* Vol. COM-34, No. 10, pp. 1043-1045, Oct. 1986.

[4] K. Yamaguchi and H. Imai, "Highly Reliable Multilevel Channel Coding System Using Binary Convolutional Codes," *Electronic Letters,* Vol. 23, No. 18, pp. 939-941, August 1987.

[5] T. Woerz and J. Hagenauer, "Multistage Coding and Decoding for a M-PSK System," *Proceedings of IEEE Global Telecommunications Conference (Globecom'90),* San Diego, Dec. 1990, pp. 501.1.1-501.1.6.

[6] A.R. Calderbank, "Multilevel Codes and Multistage Decoding," *IEEE Trans. Communications,* Vol. IT-37, No. 3, pp. 222-229, March 1989.

[7] G.J. Pottie and D.P. Taylor, "Multilevel Codes Based on Partitioning," *IEEE Trans. Information Theory,* Vol. IT-35, No. 1, pp. 87-98, Jan. 1989.

[8] J. Wu, D.J. Costello, Jr., and L.C. Perez, "On Multilevel Trellis MPSK Codes," Presented at the *1991 International Symposium on Information Theory,* Budapest, Hungary, June 1991.

[9] Y. Kofman, E. Zehavi, and S. Shamai (Shitz), "Analysis of a Multilevel Coded Modulation System," *Proceedings of the 1990 Bilkent International Conference on New Trends in Communication, Control, and Signal Processing,* Ankara, Turkey, pp. 376-382, July 1990. Also, EE Pub. No. 751 (revised version), Technion, Israel, January 1991.

[10] E. Biglieri, "Error Probability of Block-Coded Modulation," Presented at *1990 International Symposium on Information Theory and its Applications (ISITA'90),* Hawaii, U.S.A.

[11] T. Takata, S. Ujita, T. Kasami and S. Lin, "A Multi-Stage Decoding for Multi-Level Block Modulation Codes And its Error Probability Analysis," Presented at *1990 International Symposium on Information Theory and its Applications (ISITA'90),* Hawaii, U.S.A.

[12] Y. Kofman, E. Zehavi, and S. Shamai (Shitz), "A Multilevel Coded Modulation Scheme for Fading Channels," *AEU, Electronics and Communication,* Vol. 46, No. 6, pp.420-428, 1992.

[13] E. Zehavi, "8-PSK Trellis Codes on Rayleigh Channels," *IEEE Trans. Communications,* Vol. 40, No. 5, pp. 873-884, May 1992.

[14] N. Seshadri and C.W. Sundberg, "Multi-Level Trellis Coded Modulations with Large Time Diversity for the Rayleigh Fading Channel," *Proceedings of the 1990 Conference on Information Sciences and Systems,* Princeton, New Jersey, pp. 853-857.

[15] F.A. Taubin and A.N. Trofimov, "Pipeline Decoding of Embedded Trellis Codes: Error-Tolerance Analysis," *Probl. Inform. Transmission,* pp. 332-343, 1990 (English translation).

[16] D. Haccoun and G. Begin, "High-Rate Punctured Convolutional Codes for Viterbi and Sequential Decoding," *IEEE Trans. Communications,* Vol. COM-37,No. 11, pp. 1113-1125, Nov. 1989.

[17] A.J. Viterbi and J.K. Omura, *Principles of Digital Communications and Coding,* McGraw-Hill, Inc. 1979.

[18] J. Conan, "The Weight Spectra of Some Short Low-Rate Convolutional Codes," *IEEE Trans. Communications,* Vol. COM-32, No. 9, pp. 1050-1053, Sept. 1984.

[19] Y. Yasuda, K. Kashiki, and Y. Hirata, "High-Rate Punctured Convolutional Codes for Soft Decision Viterbi Decoding," *IEEE Trans. Communications,* Vol. COM-32, No. 3, pp. 315-319, March 1984.

[20] A.J. Viterbi, J.K. Wolf, E. Zehavi, and R. Padovani, "A Pragmatic Approach to Trellis-Coded Modulation," *IEEE Communications Magazine,* pp. 11-19, July 1989.

[21] R.G. Gallager, *Information Theory and Reliable Communication,* John Wiley and Sons, Inc. NY, 1968.

[22] R.G. Gallager, "A Perspective on Multiaccess Channels," *IEEE Trans. Information Theory,* Vol. 31, No. 2, pp. 124-142, March 1985.

[23] K. Yamaguchi and H. Imai, "A Study on Imai-Hirakawa Trellis-Coded Modulation Schemes," *Lecture Notes in Computer Science,* No 357, pp. 443-453, 1989, (an issue on Applied Algebra, Algebraic Algorithm and Error-Correcting Codes).

[24] J.G. Proakis, *Digital Communications,* McGraw-Hill Book Company, 1983.

Yosef Kofman was born in Tel-Aviv, Israel, on June 13, 1959. He received the B.Sc., M.Sc., and D.Sc. degrees in electrical engineering from the Technion - Israel Institute of Technology, Haifa, Israel, in 1985, 1988, and 1992, respectively.

Since 1992 he is with Teledata Communication Ltd., Herzlia, Israel, where he is involved in research and development of High-bit-rate Digital Subscriber Line-based products.

Ephraim Zehavi was born in Afula, Israel, spetember 13, 1951. He received the B.Sc. and the M.Sc. degrees in electrical engineering from the Technion - Israel Institute of Technology, Haifa, Israel, in 1977 and 1981, respectively, and the Ph.D. degree in electrical engineering from the university of Massachusetts, Amherst, in 1986.

From 1977 to 1983, he was an R&D engineer and group leader at the Department of Communication, Rafael, Armament Development Authority, Haifa, Israel. From 1983 to 1985, he was a reaserch assistant in the Department of Electrical and Computer Engineering, University of Massachesetts, Amherst. In 1985 he joined Qualcomm, Inc., San Diego, California as a senior Engineer, where he was involved in the design and development of satellite communication systems and VLSI design of Viterbi decoder chips. From 1988 to 1992 he was with the Department of Electrical Engineering, Technion, Israel. Since 1992 he is again with Qualcomm, Inc.

His current research interests include satellite communication, digital cellular telephone systems, combined coding and modulation, and personal communication systems.

Shlomo Shamai (Shitz) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion - Israel Institute of Technology, Haifa, Israel in 1975, 1981, and 1986, respectively.

During 1975-1985 he was with the Signal Corps Research Labs (Israel Defence Forces) in the capacity of a senior research engineer. Since 1986 he is with the Depertment of Electrical Engineering, Technion, Israel, where he is now an Associate Professor.

His research interests include topics in information theory and analog and digital communications. He is especially interested in theoretical limits in communication with practical constraints, digital communication in optical channels, information theoretic models for magnetic recording, channel coding, combined modulation and coding, digital spectrally efficient modulation methods, FM threshold extending techniques, and representation of signals with partial information.