On the Undetected Error Probability of Linear Block Codes on Channels with Memory

A. Kuznetsov, Francis Swarts, *Member, IEEE*,
A. J. Han Vinck, and Hendrik C. Ferreira, *Member, IEEE*

***Abstract*— In the first part of the correspondence we derive an upper bound on the undetected error probability of binary (n, k) block codes used on channels with memory described by Markov distributions. This bound is a generalization of the bound presented by Kasami *et al.* for the binary symmetric channel, and is given as an average value of some function of the composition of the state sequence of the channel. It can be extended in particular cases of Markov-type channels. As an example, such an extended bound is given for the Gilbert–Elliott channel and Markov channels with deterministic errors determined by the state. In the second part we develop a recursive technique for the exact calculation of the undetected error probability of an arbitrary linear block code used on a Markov-type channel. This technique is based on the trellis representation of block codes described by Wolf. Results of some computations are presented.**

***Index Terms*— Undetected error probability, binary codes, channel models, Gilbert–Elliott, Fritchman.**

I. Introduction

Different types of automatic repeat request systems are applied effectively in data communication systems to provide the necessary reliability in the transmission of information. The key element of all such systems is usually some block code C used solely for the detection of errors arising during the transmission of codewords through the channel. The undetected error probability $P_e(C)$ is one of the major parameters which determines the throughput efficiency. For the binary symmetric channel (BSC) with a bit error probability ϵ, the following lower and upper bounds on the minimum of $P_e(C)$, which we denote by $P(n, M, \epsilon)$, over block codes C of length n with M codewords were presented by Korzhik [3] in the mid-1960

$$P(n, M, \epsilon) \geq (M-1)\epsilon^{\overline{d}(n,M)}(1-\epsilon)^{n-\overline{d}(n,M)} \tag{1}$$

$$P(n, M=2^k, \epsilon) \leq \frac{2^k-1}{2^n-1}(1-(1-\epsilon)^n) \leq 2^{k-n} \tag{2}$$

where $\overline{d}(n, M) = nM/2(M-1)$. The lower bound was derived by applying Jensen's inequality to the expression defining $P_e(C)$ from the distance spectrum of the code C. The upper bound was obtained by averaging $P_e(C)$ over an ensemble of all binary (n, k)-codes. Using the MacWilliams identity, Leont'ev [4] proved that for any linear (n, k)-code C

$$P_e(C) \geq 2^{k-n} + (1-2\epsilon)^{n/2} - (1-\epsilon)^n \geq 2^{k-n} - (1-\epsilon)^n \tag{3}$$

Manuscript received November 8, 1993; revised April 21, 1995. The material in this correspondence was presented at the IEEE International Symposium on Information Theory, San Antonio, TX, January 1993.

A. Kuznetsov is with the Institute for Problems of Information Transmission, Moscow 101447, Russia .

F. Swarts and H. C. Ferreira are with the Laboratory for Cybernetics, Rand Afrikaans University, P.O. Box 524, Auckland Park, 2006, South Africa.

A. J. Han Vinck is with the Institute for Experimental Mathematics, University of Essen, 45326 Essen, Germany.

Publisher Item Identifier S 0018-9448(96)00026-0.

under the condition that $n/2^k \to 0,\ n \to \infty$. For

$$R = k/n \geq \log_2 2(1-\epsilon) = R_0(\epsilon)$$

and

$$k - n\log_2(1-\epsilon) \to \infty$$

inequalities (2) and (3) give the following exact asymptotic of the minimum $P_0(n, k, \epsilon)$ of $P_e(C)$ over all linear (n, k)-codes:

$$P_0(n, k, \epsilon) \sim 2^{k-n}.$$

The rate $R_0(\epsilon)$ is called the critical rate. Further significant improvements of the lower and upper bounds on $P(n, M, \epsilon)$ and $P_0(n, k, \epsilon)$ were achieved in [1], [5], [6].

At the same time, there are a number of papers dedicated to the evaluation of $P_e(C)$ for some important classes of codes on the BSC. For example, an (n, k)-code for which $P_e(C) = P_0(n, k, \epsilon)$ is called optimal for the BSC with error probability ϵ. In [3] and [4] it was shown that

1) An equidistant code is optimal on the BSC for any $\epsilon \in [0, 1/2]$.
2) A group code containing Hamming codes (in particular the Hamming code itself) is optimal on the BSC for any $\epsilon \in [0, 1/2]$.

Optimal codes constitute a very small class of all block codes. Much larger classes of codes used for error detection include the following [6]–[12]:

1) Codes that obey the bound $P_e(C) \leq 2^{k-n}$ for all $\epsilon \in [0, 1/2]$.
2) Codes with a monotonic property for which $P_e(C)$ is a nondecreasing function of $\epsilon \in [0, 1/2]$ (a subclass of the above mentioned class 1).
3) Cyclic or shortened cyclic codes.
4) Codes which are not worse than an average code from some ensemble of codes.

Although research related to the undetected error probability on the BSC is very important from the information-theoretic viewpoint, its practical value is restricted by the fact that the BSC does not always adequately describe real communication channels (which are often not memoryless at all) [13]. The Gilbert–Elliott and Fritchman channel models are examples of formal models used to analyze the characteristics of real channels with memory [14], [15], [18]. The analysis and calculation of the undetected error probability on such channels become more complicated from an analytical and computational point of view when compared with the same problem for the BSC. Few papers have been published on this subject [12].

In the next section we first give the definition of the undetected error probability $P_e(C)$ of the code C on channels with memory described by Markov distributions and consider some of its properties which are used later. In Section III we derive an upper bound on $P_e(C)$. This bound is a generalization of the bound presented in [1] for the binary symmetric channel (BSC), and is given as an average value of some function of the composition of the state sequence of the channel. It can be extended in particular cases of Markov-type channels. As an example we give such an extended bound for the Gilbert–Elliott channel. In Section IV we develop a recursive technique for the exact calculation of the undetected error probability of an arbitrary linear block code used on Markov-type channels.

This technique is based on the trellis representation of block codes described in [2] and some results of the numerical computations are presented.

II. Definition and Some Properties of the Undetected Error Probability

Let E^n be the set of all binary words of length n and C be a linear subspace of E^n of dimension k. The set C will be called a binary linear block (n,k)-code. In this correspondence we consider the transmission of words from C, called codewords or code vectors, through channels with memory described by *Markov* distributions. The function of such a channel during the transmission of n symbols is determined by the sequence $\overline{s} = (s_0, s_1, \cdots, s_n)$ of $n+1$-dependent random states s_i, $i = 0, 1, \cdots, n$, taking values in a given finite set S. Without loss of generality we will assume that $S = \{1, 2, \cdots, L\}$. The interdependence of states is defined by the following expression for the probability $P(\overline{s})$ of the state sequence $\overline{s}$:

$$P(\overline{s}) = p(s_0) \prod_{i=1}^{n} p(s_i | s_{i-1}) \tag{4}$$

where $p(s_0)$ is the probability of the initial state s_0, and $p(i|j)$ is the probability of transition from state j to state i, with

$$0 \le p(i|j) \le 1, \quad 1 \le i, j \le L$$

and

$$\sum_{i=1}^{L} p(i|j) = 1, \quad \text{for } j = 1, 2, \cdots, L.$$

An output word $\overline{y} = (y_1, y_2, \cdots, y_n) \in E^n$ of the channel can always be represented as a componentwise sum of the transmitted code word $\overline{x} = (x_1, x_2, \cdots, x_n) \in C$ and an error vector $\overline{e} = (e_1, e_2, \cdots, e_n) \in E^n$, i.e., $y_i = x_i \oplus e_i$, where $\oplus$ is binary modulo-2 addition and $1 \le i \le n$ (the operator $\oplus$ is also used to denote the modulo-2 sum of binary words, for example, $\overline{y} = \overline{x} \oplus \overline{e}$). In this correspondence we will suppose that for a given state sequence $\overline{s} \in S^{n+1}$ and a codeword $\overline{x} \in C$, the conditional distribution $P(\overline{y}|\overline{x}, \overline{s})$ on $\overline{y} \in E^n$ is defined by $\overline{s}$ and $\overline{e} = \overline{x} \oplus \overline{y}$ as the following product:

$$P(\overline{y}|\overline{x}, \overline{s}) \triangleq \prod_{i=1}^{n} \epsilon(e_i|s_i) \triangleq Q(\overline{e}|\overline{s}) \tag{5}$$

where $\epsilon(1|j) = \epsilon_j$ and $\epsilon(0|j) = 1 - \epsilon_j$ are the probabilities of erroneous and correct transmission of a single bit when the channel is in the state j, $1 \le j \le L$. Channels of the type described above will be called *channels with conditionally independent errors* (CCIE). If $P(\overline{s})$ is defined by (4), then the CCIE is called a Markov CCIE or simply MCCIE.

When a block code C is used exclusively for error detection, the received word $\overline{y}$ is decoded to

1) the information bits associated with $\overline{y}$ if $\overline{y} \in C$,
2) an error flag if $\overline{y} \notin C$.

In communication systems with this decoding rule the transmission will be erroneous if and only if $\overline{y} \ne \overline{x}$ and $\overline{y} \in C$. The probability of such an error event is the following sum:

$$P_e(C) = \sum_{\overline{s} \in S^{n+1}} \sum_{\overline{x} \in C} \sum_{\overline{y} \in C,\ \overline{y} \ne \overline{x}} P(\overline{s}) Q(\overline{x}) Pr(\overline{y}|\overline{x}, \overline{s}) \tag{6}$$

where $P(\overline{s})$, $P(\overline{y}|\overline{x}, \overline{s})$ are defined by (4) and (5), respectively, and $\{Q(\overline{x}),\ \overline{x} \in C\}$ is some probability distribution on codewords. It is not difficult to verify that under the assumption (5)

$$P_e(C) = \sum_{\overline{e} \in C,\ \overline{e} \ne 0} \sum_{\overline{s} \in S^{n+1}} P(\overline{s}) Q(\overline{e}|\overline{s}) \tag{7}$$

and this probability does not depend on the distribution $\{Q(\overline{x})\}$.

For an MCCIE the internal sum in (7) can be calculated recursively. In fact, for a given word $\overline{e} = (e_1, e_2, \ldots, e_n) \in C \subset E^n$ we have

$$\sum_{\overline{s} \in S^{n+1}} P(\overline{s}) Q(\overline{e}|\overline{s}) = \sum_{l=1}^{L} S(n, l) \tag{8}$$

where l denotes the terminating state with

$$\begin{aligned} S(n,l) &\triangleq \sum_{\overline{s} = (\overline{s'}, l),\ \overline{s'} \in S^n} P(\overline{s}) Q(\overline{e}|\overline{s}) \\ &= \epsilon(e_n | s_n = l) \sum_{j=1}^{L} p(l|j) S(n-1, j) \end{aligned} \tag{9}$$

and $S(n-1, l)$ is calculated for the first $n-1$ components of $\overline{e}$. Using matrix notation from (8) and (9) we have the following recursion:

$$\overline{S}(n) \triangleq (S(n,1), S(n,2), \cdots, S(n,L)) = \overline{S}(n-1) P A(e_n) \tag{10}$$

where P is the state transition matrix, $S(0,l) = p(l), 1 \le l \le L$, and $A(0)$, $A(1)$ are $L \times L$ matrices. In the case of $A(0)$, $a_{ij} = 1 - \epsilon_i$, for $i = j$, $i, j = 1, 2, \cdots, L$ and for $A(1)$, $a_{ij} = \epsilon_i$, for $i = j$, $i, j = 1, 2, \cdots, L$. For both $A(0)$ and $A(1)$ $a_{ij} = 0$, for $i \ne j$, $i, j = 1, 2, \cdots, L$.

Although recursion (10) simplifies the calculation of the internal sum (8) for a given $\overline{e}$, the summation of these values over all nonzero codewords $\overline{e}$ still has to be performed for the calculation of $P_e(C)$. This procedure can also be simplified by using a trellis representation of block codes suggested in [2], and will be considered in Section IV.

At the same time there is a special class of Markov channels with conditionally independent errors, containing the BSC as a particular case, for which the calculation of $P_e(C)$ can be performed exactly as for the BSC. These are so-called symmetrical channels for which all L average error probabilities

$$\sigma(e_n, i) = \sum_{j=1}^{L} \epsilon(e_n|j) p(j|i), \quad 1 \le i \le L \tag{11}$$

are equal, i.e., $\sigma(e_n, i) = \sigma(e_n)$ for all $1 \le i \le L$ and $\sigma(e_n)$ does not depend on i $(0 \le \sigma(e_n) \le 1, e_n = 0, 1)$. It is not difficult to verify that $\sigma(0) + \sigma(1) = 1$. In this case from (8)–(10) we have

$$\sum_{l=1}^{L} S(n,l) = \sigma(e_n) \sum_{l=1}^{n} S(n-1, l) = \sigma(0)^{n - w(\overline{e})} \sigma(1)^{w(\overline{e})}$$

where $w(\overline{e})$ is the Hamming weight of the word $\overline{e}$, i.e., the number of nonzero components of $\overline{e}$. From this equality and (7) it follows that for symmetrical Markov channels with conditionally independent errors

$$P_e(C) = \sum_{i=1}^{n} A_i \epsilon^i (1 - \epsilon)^{n-i} \tag{12}$$

where $\epsilon = \sigma(1)$ and A_i is the number of codewords $\overline{e} \in C$, such that $w(\overline{e}) = i$, $0 \le i \le n$. The set of $n+1$ integers A_i is called the weight spectrum of the linear code C. We formulate this result as

Theorem 1: For a symmetrical Markov channel with conditionally independent errors the undetected error probability $P_e(C)$ is defined by the spectrum of the linear (n,k)-code C and can be evaluated by (12).

Two particular cases of MCCIE will be considered further on in this correspondence. These are the Gilbert–Elliott Channel (GEC) and the Fritchman channel.

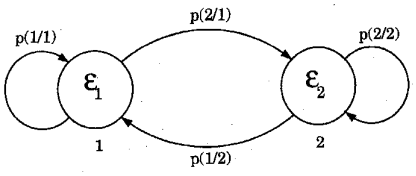

Fig. 1. State transition diagram for the Gilbert–Elliott channel model.

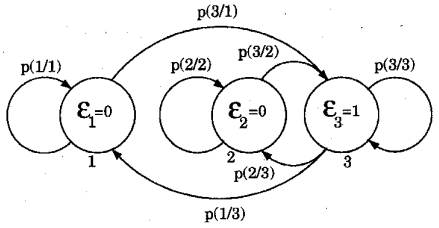

Fig. 2. State transition diagram for the Fritchman channel model.

The GEC: It has two states, i.e., $S = \{1, 2\}$, and is defined by the following four parameters: $p = p(2|1)$, $q = p(1|2)$, $\epsilon_1 = \epsilon(1|1)$, and $\epsilon_2 = \epsilon(1|2)$. Therefore, for the GEC

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$
$$A(0) = \begin{bmatrix} 1-\epsilon_1 & 0 \\ 0 & 1-\epsilon_2 \end{bmatrix}$$
$$A(1) = \begin{bmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{bmatrix}.$$

The state transition diagram of the GEC is shown in Fig. 1. We will suppose that $\epsilon_1 \leq \epsilon_2$ and, therefore, we may consider State 1 as the good state, and State 2 as the bad state of the channel. The GEC is symmetrical if $\epsilon_1(1-p) + \epsilon_2 p = \epsilon_1 q + \epsilon_2(1-q)$. This condition is equivalent to $(\epsilon_1 - \epsilon_2)(1-q-p) = 0$, and is satisfied if and only if $\epsilon_1 = \epsilon_2$ or $q = 1-p$. In the first case, when $\epsilon_1 = \epsilon_2$, the bit error probability does not depend on the state of the channel and therefore the channel is in fact the BSC. In the second case, when $q = 1-p$, the bit error probability depends on the state of the channel, but the undetected error probability $P_e(C)$ can be calculated by (12) exactly as for a BSC with bit error probability

$$\epsilon = \epsilon_1(1-p) + \epsilon_2 p = \epsilon_1 q + \epsilon_2(1-q).$$

For a given set of valid parameters p, q, ϵ_1 and ϵ_2 an interesting problem is the evaluation of the minimum $P(n, k, p, q, \epsilon_1, \epsilon_2)$ of $P_e(C)$ over all binary linear codes C of length n with k information bits. As we have already shown for $\epsilon_1 = \epsilon_2$ or $q = 1-p$ it is a well-known problem considered in [1], [3]–[11] among others. In the next section we derive an upper bound on the undetected error probability for the GEC with arbitrary valid $p, q, \epsilon_1, \epsilon_2, k$ and sufficiently large n.

Markov channel with deterministic errors: It is defined by the condition that all bit error probabilities $\epsilon(e|s)$ are equal to either 0 or 1. A widely used example of such a channel is the partitioned finite-state Fritchman channel model [18]. In the case of the Fritchman channel model the set of L states are partitioned into the subset S_g of $L' < L$ good states s with $\epsilon(1|s) = 0$, $s \in S_g$, and a subset S_b of $L - L'$ bad states s with $\epsilon(1|s) = 1$, $s \in S_b$; transitions within these subsets are prohibited. A state transition diagram for the Fritchman channel with two good states and one bad state is shown in Fig. 2 and is used as an example later ($S_g = \{1, 2\}$, $S_b = \{3\}$).

III. Upper Bounds on the Undetected Error Probability

First we derive an upper bound on the undetected error probability which is valid for an arbitrary channel with conditionally independent errors. This bound is given by the following theorem as an average value of some function of the composition $\overline{\nu(s')} = (\nu(1), \nu(2), \cdots, \nu(L))$ of the state sequence $\overline{s'}$; by definition, $\nu(i)$ is the number of components in $\overline{s'} = (s_1, s_2, \cdots, s_n)$ which are equal to i, $1 \leq i \leq L$. Note that $\overline{s'}$ is a state sequence $\overline{s}$ without the first component or initial state.

Theorem 2: For an arbitrary channel with conditionally independent errors, there exists a linear even weight (n, k)-code C for which the undetected error probability $P_e(C)$ defined by (7) satisfies the following inequality:

$$P_e(C) \leq \frac{2^k - 1}{2^n - 2} \sum_{\overline{s} \in S^{n+1}} P(\overline{s}) f(\overline{\nu(s')}) \tag{13}$$

where

$$f(\overline{\nu}) = 1 + \prod_{i=1}^{L} (1 - 2\epsilon_i)^{\nu(i)} - 2 \prod_{i=1}^{L} (1 - \epsilon_i)^{\nu(i)}. \tag{14}$$

Proof: For a fixed code C used on the CCIE the definition of $P_e(C)$ was given by (7) in Section II. To prove Theorem 2 we average $P_e(C)$ over an ensemble $\{C\}$ of all binary linear (n, k)-codes C with codewords of even Hamming weight, as it was done for the BSC by Kasami, Kløve, and Lin [1]. The number of codes in $\{C\}$ is

$$\begin{bmatrix} n-1 \\ k \end{bmatrix} = \frac{(2^{n-1} - 1)(2^{n-2} - 1) \cdots (2^{n-k-1} - 1)}{(2^k - 1)(2^{k-1} - 1) \cdots (2 - 1)} \tag{15}$$

and the number of codes $C \in \{C\}$ containing a given nonzero even weight word is

$$\begin{bmatrix} n-2 \\ k-1 \end{bmatrix} = \frac{(2^{n-2} - 1)(2^{n-3} - 1) \cdots (2^{n-k-1} - 1)}{(2^{k-1} - 1)(2^{k-2} - 1) \cdots (2 - 1)} \tag{16}$$

(see [21, p. 699]). Therefore, the average value $\overline{P_e(C)}$ of $P_e(C)$ on CCIE is given by

$$\begin{aligned} \overline{P_e(C)} &= \frac{1}{\begin{bmatrix} n-1 \\ k \end{bmatrix}} \sum_{C \in \{C\}} \sum_{\substack{\overline{e} \in C \\ \overline{e} \neq 0}} \sum_{\overline{s} \in S^{n+1}} P(\overline{s}) P(\overline{e}|\overline{s}) \\ &= \frac{\begin{bmatrix} n-2 \\ k-1 \end{bmatrix}}{\begin{bmatrix} n-1 \\ k \end{bmatrix}} \sum_{\substack{\overline{e} \in E^n,\ \overline{e} \neq 0 \\ w(\overline{e}) \text{ even}}} \sum_{\overline{s} \in S^{n+1}} P(\overline{s}) P(\overline{e}|\overline{s}) \\ &= \frac{2^k - 1}{2^{n-1} - 1} \sum_{\overline{s} \in S^{n+1}} P(\overline{s}) \sum_{\substack{\overline{e} \in E^n,\ \overline{e} \neq 0 \\ w(\overline{e}) \text{ even}}} P(\overline{e}|\overline{s}) \end{aligned} \tag{17}$$

where $w(\overline{e})$ is the Hamming weight of $\overline{e}$.

Now we calculate the internal sum in (17) which, as we will see later, depends only on the composition of the sequence $\overline{s'}$. We will use the following notations. A given state sequence $\overline{s'}$ with composition $\overline{\nu} = (\nu(1), \nu(2), \cdots, \nu(L))$ partitions the set $Z_n = \{1, 2, \cdots, n\}$ of component positions into L subsets

$$J_i = \{j \in Z_n | s_j = i\}$$

such that

$$\bigcup_{i=1}^{L} J_i = Z_n, |J_i| = \nu(i)$$

and $1 \leq i \leq L$. These subsets partition a word $\overline{e} \in E^n$ into L subwords

$$\overline{x_i} = (x_{i,1}, x_{i,2}, \cdots, x_{i,\nu(i)}) \in E^{\nu(i)}$$

with components

$$x_{i,l} = e_{j(i,l)}, \quad 1 \le l \le \nu(i) \tag{18}$$

where $j(i,1), j(i,2), \cdots, j(i,\nu(i))$ are integers of the set J_i, $1 \le i \le L$. A reverse operation for such a partition of $\overline{e} \in E^n$ will be the merging $\overline{e} = (\overline{x_1} \diamond \overline{x_2} \diamond \cdots \diamond \overline{x_L}) \in E^n$ of the words $\overline{x_i} \in E^{\nu(i)}$ which is defined by (18) and depends on the state sequence $\overline{s'} \in S^n$. Finally, let us denote the subsets of E^n containing the words of even and odd weight by $E^{n,0}$ and $E^{n,1}$, respectively; $E^{n,0} \cup E^{n,1} = E^n$.

Using these notations and (18), we can rewrite the internal sum in (17) as follows:

$$\sum_{\substack{\overline{e} \in E^n,\ \overline{e} \ne 0 \\ w(\overline{e}) \text{ even}}} P(\overline{e}|\overline{s}) = \sum_{\overline{\tau}} S_{\overline{\tau}} \tag{19}$$

where the sum is taken over all even weight binary words $\overline{\tau} = (\tau(1), \tau(2), \cdots, \tau(L)) \in E^L$, and

$$\begin{aligned} S_{\overline{\tau}} &= \sum_{\substack{\overline{e} = (\overline{x_1} \diamond \overline{x_2} \diamond \ldots \diamond \overline{x_L}), \\ \overline{x}_i \in E^{\nu(i), \tau(i)}}} P(\overline{e}|\overline{s}) \\ &= \prod_{i=1}^{L} \left[\sum_{\substack{0 \le j \le \nu(i) \\ j = \tau_i \bmod 2}} \binom{\nu(i)}{j} \epsilon_i^j (1-\epsilon_i)^{\nu(i)-j} \right]. \end{aligned} \tag{20}$$

Since for arbitrary numbers a and b, such that $a + b = 1$,

$$\sum_{\substack{0 \le i \le n \\ i \text{ even}}} \binom{n}{i} a^{n-i} b^i = \frac{1}{2}(1 + (a-b)^n)$$

$$\sum_{\substack{0 \le i \le n \\ i \text{ odd}}} \binom{n}{i} a^{n-i} b^i = \frac{1}{2}(1 - (a-b)^n)$$

then

$$S_{\overline{\tau}} = 2^{-L} \prod_{i=1}^{L} (1 + (-1)^{\tau(i)} (1 - 2\epsilon_i)^{\nu(i)}). \tag{21}$$

Using the equality

$$\sum_{\substack{\overline{\tau} \in E^L \\ w(\overline{\tau}) \text{ even}}} \prod_{i=1}^{L} (1 + (-1)^{\tau(i)} a_i) = \left(1 + \prod_{i=1}^{L} a_i\right) 2^{L-1}$$

which is valid for arbitrary real numbers $a_1, a_2, \cdots, a_n$, from (19) and (21) we have

$$\sum_{\substack{\overline{e} \in E^n,\ \overline{e} \ne 0 \\ w(\overline{e}) \text{ even}}} P(\overline{e}|\overline{s}) = 2^{-1} \left[1 + \prod_{i=1}^{L} (1 - 2\epsilon_i)^{\nu(i)} - 2 \prod_{i=1}^{L} (1 - \epsilon_i)^{\nu(i)} \right]. \tag{22}$$

Therefore, as it follows from (17) and (22)

$$\overline{P_e(C)} = \frac{2^k - 1}{2^n - 2} \sum_{\overline{s} \in S^{n+1}} P(\overline{s}) f(\overline{\nu}(\overline{s'})) \tag{23}$$

where $f(x)$ is defined by (14). Since at least one code C in the ensemble $\{C\}$ has $P_e(C) \le \overline{P_e(C)}$, then from (23) we come to the conclusion formulated earlier in Theorem 2.

For many particular cases of channels with conditionally independent errors the upper bound (13) can be extended to get simpler bounds. First we illustrate this idea for the GEC.

Gilbert–Elliott channel: In this case

$$f(\overline{\nu}) = 1 + (1-2\epsilon_1)^{n-x}(1-2\epsilon_2)^x - 2(1-\epsilon_1)^{n-x}(1-\epsilon_2)^x$$

where $x = \nu(2)$ is the number of "bad" States 2 in the state sequence $\overline{s'}$. To extend the bound (13) we first show that $f(\overline{\nu})$ is a convex $\cap$ function of x for sufficiently large n. In fact, $f(\overline{\nu}) = 1 + 2(1-\epsilon_1)^n \phi(x)$, where

$$\phi(x) = A\beta^x - \alpha^x$$

$$A = 2^{-1}((1-2\epsilon_1)/(1-\epsilon_1))^n$$

$$0 \le \beta = (1-2\epsilon_2)/(1-2\epsilon_1) \le \alpha = (1-\epsilon_2)/(1-\epsilon_1) \le 1.$$

Therefore, $f(\overline{\nu})$ is a convex $\cap$ function of x, if and only if

$$\frac{d^2\phi}{dx^2} = A\beta^x (\ln\beta)^2 - \alpha^x (\ln\alpha)^2 \le 0.$$

This inequality is equivalent to the following one:

$$A\left(\frac{\beta}{\alpha}\right)^x \left(\frac{\ln\beta}{\ln\alpha}\right)^2 \le 1. \tag{24}$$

Here we note that if (24) is valid for $x = 0$, then it is valid for any $x > 0$ as well. But for $x = 0$ it can be rewritten as follows:

$$n \ln\left(\frac{1-\epsilon_1}{1-2\epsilon_1}\right) \ge -\ln 2 + 2\ln\frac{\ln((1-2\epsilon_2)/(1-2\epsilon_1))}{\ln((1-\epsilon_2)/(1-\epsilon_1))}. \tag{25}$$

Therefore, $f(\overline{\nu})$ is a convex $\cap$ function of $x = \nu(2)$ for n satisfying (25), and we can apply Jensen's inequality to upper-bound the right side of (23). This gives us the following theorem.

Theorem 3: For a given set of four parameters p, q, ϵ_1 and ϵ_2 of GEC ($0 \le p, q \le 1$, $0 < \epsilon_1 \le \epsilon_2 \le 0.5$), there exists a linear binary even weight (n,k)-code C of length n satisfying (25), for which the undetected error probability $P_e(C)$ satisfies the following inequality:

$$\begin{aligned} P_e(C) &\le F(n,k,\epsilon_1,\epsilon_2,\nu) \\ &\triangleq \frac{2^k - 1}{2^n - 2}[1 + (1-2\epsilon_1)^{n-\nu}(1-2\epsilon_2)^\nu \\ &\qquad - 2(1-\epsilon_1)^{n-\nu}(1-\epsilon_2)^\nu] \end{aligned} \tag{26}$$

where

$$\nu = \sum_{\overline{s} \in S^{n+1}} \nu(2) P(\overline{s}) \tag{27}$$

is an average number of "bad" States 2 in the sequence $\overline{s'} = (s_1, s_2, \cdots, s_n)$.

Let $P(n,k,p,q,\epsilon_1,\epsilon_2)$ be the minimum of $P_e(C)$ on the GEC over all binary linear (n,k) codes. Then we have

Corollary 1: For an arbitrary set of valid parameters $n, k, p, q, \epsilon, \delta$

$$P(n,k,p,q,\epsilon_1,\epsilon_2) \le F(n,k,\epsilon_1,\epsilon_2,\nu)$$

where $F(n,k,\epsilon_1,\epsilon_2,\nu)$ and ν are defined by (26) and (27), respectively.

Corollary 2: If the initial probability distribution $\overline{p} = (p(1), p(2))$ on s_0 is the stationary distribution of the Markov chain with the transition matrix P, i.e., $\overline{p} = P\overline{p}$, and n satisfies (25), then

$$P(n,k,p,q,\epsilon_1,\epsilon_2) \le F(n,k,\epsilon_1,\epsilon_2,nq/(p+q)). \tag{28}$$

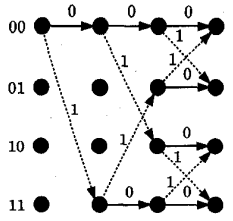

Fig. 3. Trellis for the $(3, 1)$-code described in the text.

Markov channels with deterministic errors: Referring to the description of these channels in Section II and the definition (14) of the function f, we get that in this case

$$f(\overline{v}) = 1 + (-1)^w - 2a$$

where w is the number of "bad" states s with $\epsilon(1|s) = 1$ in the state sequence $\overline{s'}$; $a = 0$ if $w > 0$, $a = 1$ if $w = 0$. Therefore, as a corollary of Theorem 2, we have

Theorem 4: For a Markov channel with deterministic errors there exists a linear even-weight (n, k)-code C for which

$$P_e(C) \leq P\{w \text{ is even}\} - P\{w \text{ is odd}\} - 2P\{w = 0\}.$$

IV. Recursive Calculation of the Undetected Error Probability

In this section we present a method for the recursive calculation of the undetected error probability of the (n, k)-code C given by its parity check matrix $H = [\overline{h_1}\ \overline{h_2} \cdots \overline{h_n}]$, where $\overline{h_i} \in E^{n-k}$ is the ith column of H, $1 \leq i \leq n$. This method is based upon the trellis representation of linear codes described by Wolf [2].

The trellis of the code C is a directed graph with $n + 1$ sets of nodes I_j, $0 \leq j \leq n$. The nodes from I_j are said to be at depth j, $0 \leq j \leq n$, and are identified (labeled) by binary words of length $r = n - k$. A binary word identifying the node $x \in I_j$ is called its label. We should note that not all words from E^r may correspond to the nodes in the set I_j. The edges of the trellis connect only pairs of nodes (x, y) such that $x \in I_j$ and $y \in I_{j+1}$ for some $j = 0, 1, \cdots, n-1$ (edges are oriented from x to y). The number of nodes at each depth and all interconnections of nodes in the trellis is defined by the following two rules.

1) At depth 0 the trellis contains only one node labeled by the all-zero word $\overline{0} \in E^r$.
2) For each $j = 0, 1, \cdots, n-1$, the node x with the label $\overline{\alpha}$ at depth j is connected to the node y with the label $\overline{\beta}$ at depth $j + 1$ if and only if

$$\overline{\beta} = \overline{\alpha} + e\overline{h_{j+1}} \tag{29}$$

for some $e = 0, 1$. The edge connecting nodes x and y is labeled by the binary symbol e defined by (29).

The trellis constructed according to these rules for the $(3, 1)$-code with the parity-check matrix

$$H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

is shown in Fig. 3.

As we can see from the description of the trellis for any binary word $\overline{e} = (e_1, e_2, \cdots, e_n) \in E^n$ there is a path on the trellis which originates in the single node at depth 0 and terminates at some node at depth n. It is not difficult to verify using [2] that the trellis has the following two properties.

1) All paths corresponding to codewords start at depth 0 and terminate at depth n in the node labeled by $\overline{0} \in E^r$.
2) The paths corresponding to different codewords diverge at some node, i.e., they are different.

Now we show how the trellis described above can be used for the recursive calculation of the undetected error probability $P_e(C)$. In

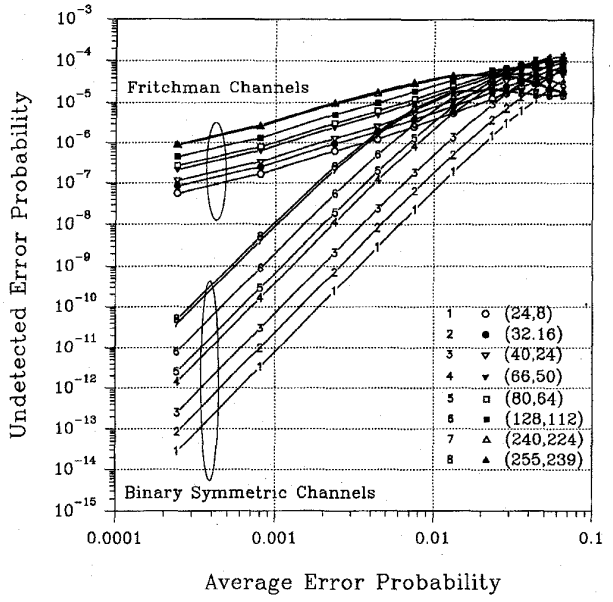

Fig. 4. Comparison of $P_e(C)$ for various CRC-16 codes on the three-state Fritchman channels described in [17] as well as on the BSC.

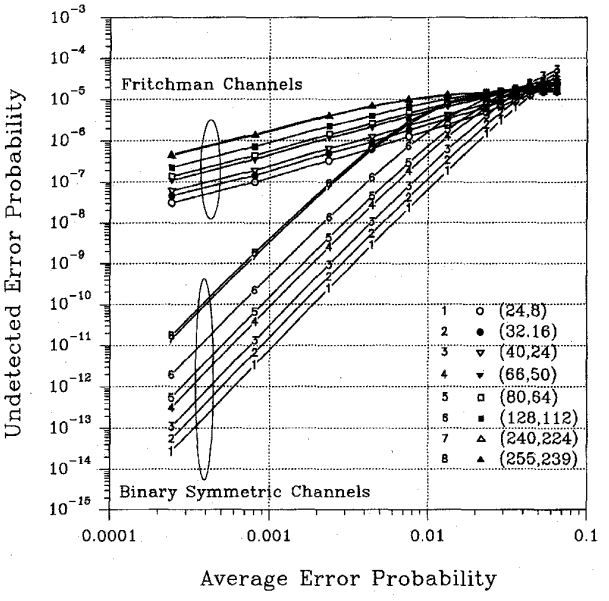

Fig. 5. Comparison of $P_e(C)$ for various CRC-CCITT codes on the three-state Fritchman channels described in [17] as well as on the BSC.

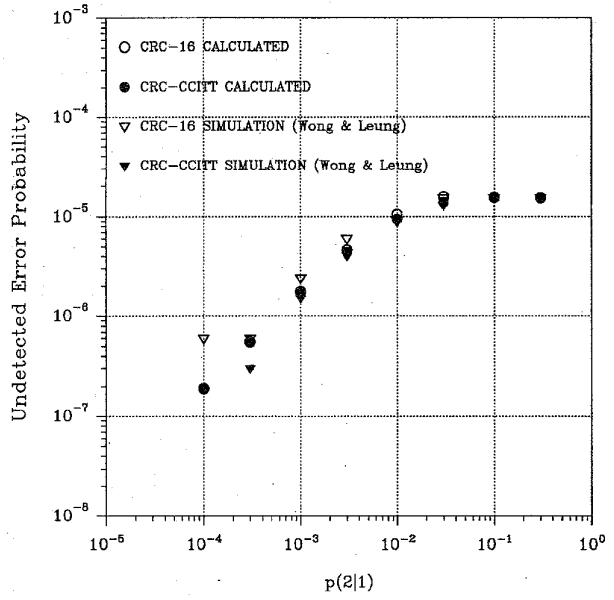

Fig. 6. Comparison of Wong and Leung's simulation results [19] to our direct calculation of $P_e(C)$. $\epsilon_1 = 0$, $\epsilon_2 = 0.5$, and $p(1|2) = 0.01$.

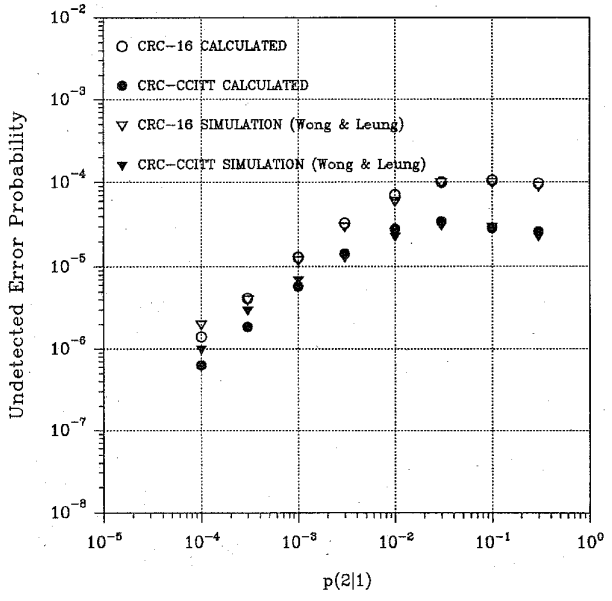

Fig. 7. Comparison of Wong and Leung's simulation results [19] to our direct calculation of $P_e(C)$. $\epsilon_1 = 0$, $\epsilon_2 = 0.1$, and $p(1|2) = 0.01$.

Section II we obtained the following expression (see (7) and (8)):

$$P_e(C) = \sum_{\overline{e} \in C, \overline{e} \neq 0} \sum_{l=1}^{L} S(n,l) \tag{30}$$

where $S(n,1), S(n,2), \cdots, S(n,l)$ are the components of the vector $\overline{S}(n)$ calculated recursively by (10) for a given binary word $\overline{e} = (e_1, e_2, \cdots, e_n)$. There is a trivial method of doing all of the summations in (30): first we can calculate internal sums in (30) for all nonzero codewords $\overline{e}$, then we sum the results of the previous step.

This method can be used when the number of codewords is not too large, e.g., for codes with low rates $R = k/n$. Some important

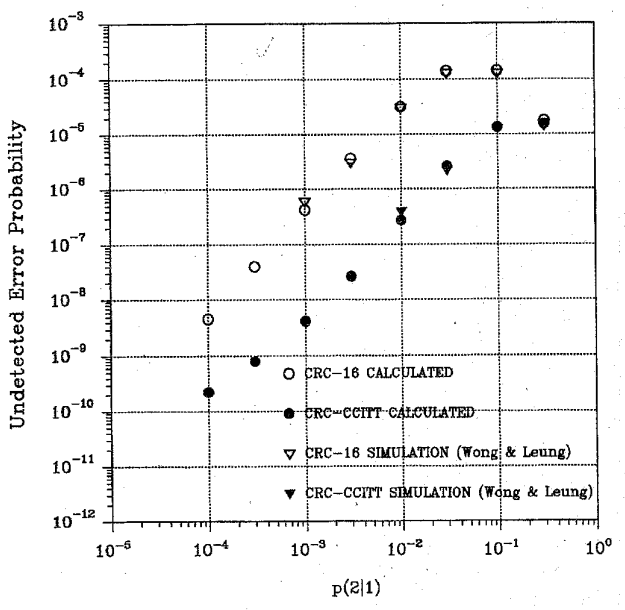

Fig. 8. Comparison of Wong and Leung's simulation results [19] to our direct calculation of $P_e(C)$. $\epsilon_1 = 0$, $\epsilon_2 = 0.5$, and $p(1|2) = 0.3$.

codes have high rates and relatively small redundancies $r = n - k$. In the last case, the following procedure is more efficient from a computational point of view. Referring to the trellis described above and recursion (10) from Section II, we can state this procedure as follows. For each node x at depth j, $0 \leq j \leq n$, we assign the vector $\overline{S}(j,x)$ with real components in accordance with the following rules:

1) For $j = 0$ let $\overline{S}(0,\overline{0})$ be the probability distribution $(p(0), p(1), \cdots, p(L))$ of the initial state of the channel.
2) Suppose that the vectors $\overline{S}(j,x)$ are already assigned to all nodes x at depth $j < n$, and y is a node at depth $j+1$ connected to the nodes $x_1, x_2, \cdots, x_v$ at depth j by edges labeled by binary symbols $a_1, a_2, \cdots, a_v$, respectively. Then vector $\overline{S}(j+1,y)$ assigned to the node y at depth $j+1$ is determined by the equality

$$\overline{S}(j+1,y) = \sum_{i=1}^{v} \overline{S}(j,x_i) PA(a_i) \tag{31}$$

where P is the state transition matrix, and $A(a_i)$ are the $L \times L$ matrices defined in Section II.

Theorem 5: The undetected error probability $P_e(C)$ of the code C is equal to the sum of the components of the vector $\overline{S}(n,\overline{0})$ assigned to the node with the label $\overline{0}$ at depth n of the trellis.

The proof of this theorem follows directly from (30) and the description of the trellis. The described recursive technique was applied to the calculation of the undetected error probability for the well-known CRC-16 and CRC-CCITT codes [20] used for error detection on Gilbert–Elliott and Fritchman channels. Results of calculations are given in Figs. 4–9. In Figs. 4 and 5 we present $P_e(C)$ for various CRC-16 and CRC-CCITT codes used on Fritchman channels with state transition diagrams as shown in Fig. 2 and the BSC. The average error probability shown on these graphs represents the bit error probability for the BSC and the average bit error probability for the Fritchman channel, $\epsilon = P_3$. The parameters of the Fritchman channel used for these calculations are taken from [17], and are typical of Rayleigh fading channels.

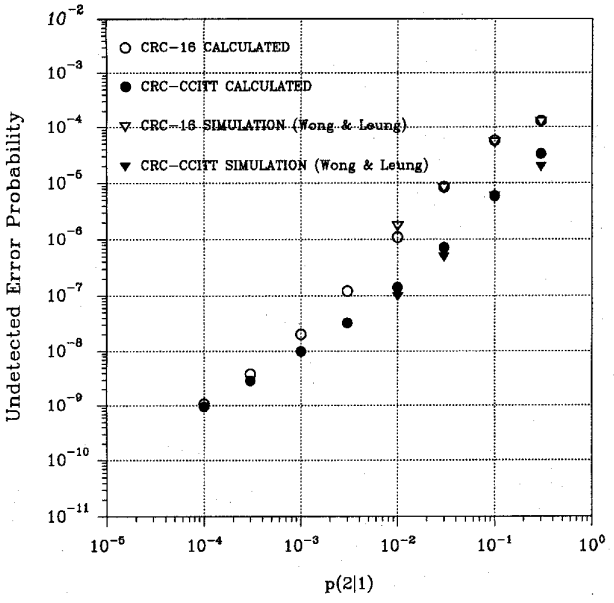

Fig. 9. Comparison of Wong and Leung's simulation results [19] to our direct calculation of $P_e(C)$. $\epsilon_1 = 0$, $\epsilon_2 = 0.1$, and $p(1|2) = 0.3$.

In Figs. 6–9 we compare the simulation results from [19] with our results of exact calculation of the undetected error probability for the CRC-CCITT and CRC-16 with $n = 66$ and $k = 50$ on Gilbert channels with the parameters as shown in Figs. 6–9. These figures clearly show simulation and calculation to give similar results. However, Wong and Leung indicated in their paper [19] that it took approximately 6.5 h of CPU time on a Sun SPARC station 1 to simulate a single set of values in the graphs. In our case it took approximately 8 h to calculate all the values in one graph for a particular code. Seeing that Wong and Leung only made use of 10^7 random words, it was impossible to find undetected error probabilities through simulation at some of the average error probabilities shown in the graphs. The technique which we have proposed, however, delivers results independent of any channel parameters, depending only on the redundancy of the code, $r = n - k$. We have therefore been able to calculate undetected error probabilities for average error probabilities which make simulation nonfeasible in determining undetected error probabilities.

A similar approach to that described in this section has been used successfully by Kittel [22] for a variety of problems in coding design.

V. Conclusions

In this correspondence we have presented a general upper bound on the undetected error probability of linear (n, k)-codes on channels with conditionally independent errors. This bound was extended to specific types of channels described by channel models such as the Gilbert–Elliott and Fritchman models.

A compact method for the exact evaluation of $P_e(C)$, based on the trellis representation of block codes, was also introduced and the effectiveness of this method was demonstrated through comparison with simulation results presented in [19].

References

[1] T. Kasami, T. Kløve, and S. Lin, "Linear block codes for error detection," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 131–136, Jan. 1983.

[2] J. K. Wolf, "Efficient maximum likelihood decoding of linear block codes using a trellis," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 76–80, 1978.

[3] V. I. Korzhik, "Bounds on undetected error probability and optimum group codes in a channel with feedback," *Radiotekhnika*, vol. 20, pp. 27–33, 1965.

[4] V. K. Leont'ev, "Coding with error detection," *Probl. Pered. Inform.*, vol. 8, pp. 6–14, 1972.

[5] V. I. Levenshtein, "Bounds on the probability of undetected error," *Probl. Pered. Inform.*, vol. 13, pp. 3–18, 1977.

[6] J. K. Wolf, A. M. Michelson, and A. H. Levesque, "On the probability of undetected error for linear block codes," *IEEE Trans. Commun.*, vol. COM-30, pp. 317–324, Feb. 1982.

[7] S. K. Leung-Yan-Cheong and M. E. Hellman, "Concerning a bound on undetcted error probability," *IEEE Trans. Inform. Theory.*, vol. IT-22, pp. 235–237, Mar. 1976.

[8] T. Kasami and S. Lin, "On the probability of undetected error for the maximum distance seperable codes," *IEEE Trans. Commun.*, vol. COM-32, pp. 998–1006, Sept. 1984.

[9] P. Perry, "Necessary conditions for good error detection," *IEEE Trans. Inform. Theory*, vol. 37, pp. 375–378, Mar. 1991.

[10] C. T. Ong and C. Leung, "On the undetcted error probability of triple-error-correcting BCH-codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 673–678, May 1991.

[11] T. Fujiwara, T. Kasami, and S. Feng, "On the monotonic property of the probability of undetected error for a shortened code," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1409–1411, May 1991.

[12] V. K. Agarwal and A. Ivanov, "Computing the probability of undetected error for shortened cyclic codes," *IEEE Trans. Commun.*, vol. 40, pp. 494–499, Mar. 1992.

[13] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their application to error control," *Proc. IEEE*, vol. 66, pp. 724–744, July 1978.

[14] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sept. 1960.

[15] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sept. 1963.

[16] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[17] F. Swarts, H. C. Ferreira, and D. R. Oosthuizen, "Renewal channel models for PSK on slowly fading Rayleigh channels," *Electron. Lett.*, vol. 25, pp. 1514–1515, Oct. 26, 1989.

[18] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 221–227, Apr. 1967.

[19] B. Wong and C. Leung, "On computing undetected error probabilities on the Gilbert channel," in *Proc. IEEE Pacific Rim Conf. on Telcommunication and Computing*, 1992, pp. 356–359.

[20] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[21] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1977.

[22] L. Kittel, "Applications of the Viterbi algorithm to error control coding design," in *Proc. 1984 Int. Zürich Sem. on Digital Communications*, pp. 75–83.