Error Control Strategies for the Wireless Channel

Michele Zorzi and Ramesh R. Rao

Center for Wireless Communications, University of California San Diego, 9500 Gilman Dr.
La Jolla, CA 92037-0407, USA – fax: +1-619-534-2486 – e-mail: {zorzi, rao}@ece.ucsd.edu

Abstract—The requirements of emerging new communications services differ in many critical ways from traditional voice applications. The effect of errors on voice transmission and techniques to overcome them are relatively well understood. On the other hand, the impact of errors on multimedia transmissions is not fully understood. The Gilbert-Elliott model is adequate for characterizing the bursty errors that occur over a communication channel. The channel “seen” by applications is the physical channel as modified by the error correcting mechanisms used at the physical level. Therefore, correlations in the residual error process are relevant to the design of encoding algorithms. In this paper we study the residual error process and characterize the block error process beyond the marginal statistics. A Markovian model for the block errors is considered and shown to be adequate. Derivations of the parameters of the block error process are presented in terms of the parameters of the bit/symbol error process. For voice transmissions the use of interleaving and a moderate amount of error correction is found to be desirable, but at higher data speeds, an effective interleaving strategy is found to require very large memory.

I. INTRODUCTION

Most communication channels are prone to errors due to various physical impairments that occur on the channel. Error correcting codes have been developed to overcome or reduce the impact of these errors. It is usually inefficient to attempt to provide a very high degree of protection and therefore some amount of residual errors pass through undetected. In applications such as voice, these residual errors simply result in speech quality degradation which is designed to be within the range of tolerance. In contrast, applications such as data or lossless image transfers must ensure that there are no residual errored bits. Now, the channel “seen” by these higher layer applications is not the raw physical channel but the physical channel as modified by the error correcting mechanisms used at the physical level. Thus, the residual error process is most relevant to the effective design of higher layer encoding algorithms for image, video and other mixed media transmission. We believe that the characterization of the residual error process is one of the key technical problems that must be resolved in order to efficiently support PCS over the wireless channel.

A complete statistical characterization of the bit/symbol error process can be complex and often only the marginal statistics are computed. At the same time, higher layer applications manage data transfer in blocks that contain multiple bits or symbols and employ various block error detection and retransmission schemes. Even without examining all of the specifics, it can be argued that the block error process is likely to be sensitive to correlations in the residual bit/symbol error process. Hence we are motivated to study the residual bit/symbol error process in order to characterize the block error process beyond the marginal statistics.

Specifically, in this paper, we consider a Gilbert-Elliott model for the channel [1, 2, 3], which captures the bursty nature of bit/symbol errors. It is assumed that the channel can be in two states: a good state, state 0, where errors occur with small probability, $P_e(0)$, and a bad state, state 1, where errors occur with high probability, $P_e(1)$. In the original model by Gilbert [1], later generalized by Elliott, $P_e(0) = 0$ was assumed.

In [3], Yee and Weldon found the decoding error probability for a Gilbert-Elliott channel, using both a recursive technique and a combinatorial approach, which provide an algorithm and a closed-form expression, respectively, for the probability that $m$ symbols out of $n$ are in error, $P(m, n)$. This quantity is useful in evaluating the performance of an error correcting schemes that can correct up to $t$ symbols errors in a block of $n$. Interleaving, which modifies the channel parameters but not its Markov character, is also considered. The authors in [3] limit themselves to computing the first-order statistics of the block error, i.e., no correlation between errors in different data blocks is considered.

These approaches are not adequate from the perspective of higher layer applications. Two problems arise when one tries to describe the channel at the block level by means of higher-order statistics. First of all, what order statistics are needed in order to adequately describe the channel behavior? Secondly, how can one compute these joint distributions?

The combinatorial analysis in [3] is rather complex, and its extension to higher-order statistics appears a hard (or at least tedious) task. On the other hand, the error process on a Gilbert-Elliott channel, being a probabilistic function of a Markov chain (with the two channel states, good and bad), can be adequately studied within the general analytical framework given in [4]. Furthermore, the comprehensive analytical framework given in [4] is quite powerful and may be used to compute the joint statistics, in principle, of any order.

We describe the Channel Model in Section II. In Section III we study the Markov approximation at the block level. Results are presented in Section IV, and Section V concludes the paper.

II. CHANNEL MODEL

The physical channel is modeled as assuming one of two states (a “good” state, 0, and a “bad” state, 1), each having an associated error probability. The transitions between these two states occur at discrete time instants, so that the channel is assumed to stay in a given state for an integer multiple of some time unit, which can be the duration of a bit, a symbol [3], or even a packet [5] (the terms “block” and “packet” will be used interchangeably throughout the paper). Let $\gamma_n = 0$ if the channel is good during the $n$-th time unit, and $\gamma_n = 1$ otherwise. Then $\gamma_n$ is a binary Markov process whose transition matrix, with elements $P[\gamma_{n+1} | \gamma_n]$, is given by

$$ P = \begin{pmatrix} P[0|0] & P[1|0] \\ P[0|1] & P[1|1] \end{pmatrix} = \begin{pmatrix} p_00 & p_01 \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} p & 1-p \\ r & 1-r \end{pmatrix}. $$(1)
The length of a burst (e.g., the number of time units the channel stays in the bad state) is a geometric random variable with mean $1/r$, and similarly for the time the channel is in good state (mean $1/(1-p)$). The steady-state probability of the channel being in bad state is given by

$$\pi_1 = \frac{1-p}{r+1-p}. \quad (2)$$

If $P_s(i)$ is the error probability given that the channel is in state $i$, we can find the steady-state error rate as [3]

$$\varepsilon = \pi_0 P_s(0) + \pi_1 P_s(1) = \frac{P_s(0)r + P_s(1)(1-p)}{r+1-p}. \quad (3)$$

We can keep track of events which are associated with transitions by “tagging” the transition diagram of the Markov chain appropriately, i.e., by labeling the edges of the chain flow graph with some transfer functions. Therefore, on the top of the two-state Markov chain which is needed to keep memory of the channel status, we can build an analytical structure to keep track of events (in particular, errors).

More specifically, let $g(n) = (g_1(n), \ldots, g_k(n))$ be a vector random process that tracks various metrics associated with the transition from $\gamma_n$ to $\gamma_{n+1}$. This implies that, given $\gamma_n = i$, $g(n)$ is independent of $g(\ell)$ and $\gamma_\ell$, for $\ell < n$, and $g(n-1)$ is independent of $g(\ell)$ and $\gamma_{\ell+1}$ for $\ell \geq n$. Let $\phi_{ij}(k|n), k = (k_1, k_2, \ldots, k_n)$ be defined as

$$\phi_{ij}(k|n) = P[\gamma_n = j, \sum_{\ell=0}^{n-1} g(\ell) = k | \gamma_0 = i]. \quad (4)$$

Since $\gamma_n$ is homogeneous, define for all $n$,

$$\Lambda_{ij}(k) \triangleq \phi_{ij}(k|1) = P[\gamma_n = j, g(n-1) = k | \gamma_{n-1} = i]. \quad (5)$$

Then, by using the total probability theorem, the fact that $\gamma_n$ is Markov and the independence properties of $g(n)$, one can write the following recursive relationship

$$\phi_{ij}(k|n) = \sum_{m=0}^{\infty} P[\gamma_{n-1} = m, \sum_{\ell=0}^{n-2} g(\ell) = k-b | \gamma_0 = i] \times P[\gamma_n = j, \sum_{\ell=0}^{n-1} g(\ell) = k-b, \gamma_{n-1} = m, \sum_{\ell=0}^{n-2} g(\ell) = k-b]$$

$$= \sum_{m=0}^{\infty} \sum_{b=0}^{\infty} \phi_{jm}(k-b|n-1) \Lambda_{mj}(b), \quad n > 0 \quad (6)$$

$$\phi_{ij}(0|0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \quad (7)$$

In general, depending on the statistics of $g(n)$, $k$ may not take arbitrary values, and this results in some values of $\phi_{ij}(k|n)$ being zero. For example, if $\kappa_\ell(n)$ is non-zero for all $\ell$, we have $\phi_{ij}(k|n) = 0$ if $k_\ell < 0$ for some $\ell$.

We note that all these recursive relationships for the probability distributions have corresponding transform relationships for the generating functions [4], which are sometimes easier to handle.

This general setup will allow us to compute the joint statistics of the block errors on a Gilbert-Elliott channel with or without interleaving. In order to do so, $g(n)$ is to be specified and the functions $\Lambda_{mj}(b)$ are to be found, and therefore the two cases are treated separately.

### A. No interleaving

In this case, successive data blocks are sent one after the other. Due to the Markov character of the channel, the functions $\phi_{ij}(k|n)$, where $k$ is the number of errors in $n$ consecutive symbols, are sufficient in order to completely describe the error process. In this case, $\kappa(n)$ is a scalar, which is set to 1 if symbol $n$ is in error, and to 0 otherwise. For this case, Eq. (5) evaluates to

$$\Lambda_{mj}(b) = p_{mj} \lambda(b, m), \quad (8)$$

where

$$\lambda(1, m) \triangleq P[\text{symbol in error } | \gamma_n = m] = P_s(m), \quad \forall n,$$

and $\lambda(0, m) = 1 - \lambda(1, m)$. Eq. (6) can be written as

$$\phi_{ij}(k|n) = \sum_{m=0}^{\infty} \left[ \phi_{im}(k|n-1)(1-P_s(m)) + \phi_{im}(k-1|n-1)P_s(m) \right] p_{mj}, \quad (10)$$

with initial conditions

$$\phi_{ij}(0|0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \quad (11)$$

To study the higher order properties of the block error process, let $N$ be the block size, and let $\phi_{i(j)}(k) \triangleq \phi_{i(j)}(k|N)$. The joint probability of having $k_1, k_2, k_3$ errors in three consecutive packets and of ending in state $\gamma_N = j$ given that $\gamma_0 = i$, can be simplified to

$$\phi_{i(j)}(k_1, k_2, k_3) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \phi_{in}^{(1)}(k_1) \phi_{mn}^{(1)}(k_2) \phi_{mj}^{(1)}(k_3). \quad (13)$$

Let $\beta_i = 1$ if the $n$-th data block is in error and $\beta_i = 0$ otherwise. Let us consider a code with error capability $t$, and let $\Gamma_0 = \{0, 1, \ldots, t\}$ and $\Gamma_1 = \{t+1, \ldots, N\}$. The distribution of $\beta$ with the channel ending in state $\gamma_N = j$ given that it started in state $\gamma_0 = i$ can be found as

$$\psi_{i}^{(1)}(\beta) = \sum_{k \in \Gamma} \phi_{i(k)}, \quad (14)$$

and the analogous quantity for three consecutive blocks is

$$\psi_{i}^{(3)}(\beta_1, \beta_2, \beta_3) = \sum_{k_1 \in \Gamma} \sum_{k_2 \in \Gamma} \sum_{k_3 \in \Gamma} \phi_{i(k_1, k_2, k_3)} \quad (15)$$

The unconditional distributions can be found by removing the condition on the initial state and by summing on the final state, to get

$$\psi = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \pi_i \psi_{ij} \quad (16)$$

Although we worked out the details for the computation of the third-order statistics, it is clear that the above approach can be extended to any order. However, as we will see, the process $\beta_i$ itself can often be very well approximated by a two-state Markov process, so that the second order statistics is adequate.
**B. Interleaving**

In the previous subsection, when interleaving was not used, it was possible to nicely decouple the error occurrences in successive packets by simply conditioning on the channel status at the packet boundary, due to the memoryless Markov property. When interleaving is used, this is no longer possible, and a more complex development is needed.

Let $d$ be the interleaving depth, and $N$ be the block length. Then, prior to transmission, the blocks are written as rows of a $d \times N$ matrix, which is read by column to obtain the data stream which is actually sent on the channel. At the receiving end, the dual operation is performed and the original order of the data is restored. This operation has the effect of mitigating the effect of the channel memory and of breaking error bursts so that error correcting codes can be more effectively used, while requiring some memory and causing some delay.

A packet is formed by symbols in positions $0, d, 2d, \ldots, (N - 1)d$, the following packet is composed of symbols in positions $1, d + 1, 2d + 1, \ldots, (N - 1)d + 1$, and so on. Let us sample the Markov process $\gamma_n$ to obtain another Markov process, $\eta_n$, and let $\kappa\ell(n) = 1$ if symbol $nd + \ell - 1$ is in error, and 0 otherwise, for $\ell = 1, 2, 3$. Also, let $\phi(1, k_1, k_2, k_3|\ell)n$ be the joint probability of having $k_1$, $k_2$ and $k_3$ errors in three consecutive packets with the channel ending in state $\eta_n = j$, given that it started in state $\eta_0 = i$.

Based on its definition, $\Lambda_{n}(b_1, b_2, b_3) = \Lambda_{n}(b_1)$ is evaluated in Eq. (17) (see top of the page), where we used the fact that, since error on symbol $\ell$ depends only on $\gamma_\ell$, $P(\kappa_\ell(n) = b_\ell|\gamma_\ell = m_\ell, i \geq 0) = \lambda(b_\ell, m_{n:d+\ell-1}),$ (18)

and where $p_{ij}$ are the entries of the channel transition matrix, $P$, and $p^{(x)}_{ij}$ are those of the $x$-step channel transition matrix, $P^x$.

Finally, (6) becomes

$$\phi_{ij}(k_1, k_2, k_3|\ell)n = \sum_{m=0}^{1} \sum_{b_1=0}^{1} \sum_{b_2=0}^{1} \Lambda_{n}(b_1, b_2, b_3) \times$$

$$\phi_{im}(k_1-b_1, k_2-b_2, k_3-b_3|n-1)$$

with initial conditions

$$\phi_{ij}(k|\ell)n = 0 \text{ if } n < 0 \text{ or } k_\ell < 0 \text{ for some } \ell$$

$$\phi_{ij}(0,0,0|0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The joint block error statistics can therefore be computed as in the previous subsection.

Note that the condition on $i$ is not necessary, and the recursive relationships could be written in terms of $\phi_i = \pi_0 \phi_{0j} + \pi_1 \phi_{1j}$, so that two recursions are required instead of four.

**III. MARKOV APPROXIMATION AT THE BLOCK LEVEL**

In order to evaluate the accuracy of a Markov approximation for the process $\beta_n$, consider the following. Let $I(\beta_n; \beta_{n-1}\beta_{n-2})$ be the average mutual information between the r.v. $\beta_n$ and the past two transmissions, $\beta_{n-1}$ and $\beta_{n-2}$. We can write [6]

$$I(\beta_n; \beta_{n-1}\beta_{n-2}) = I(\beta_n; \beta_{n-1}) + I(\beta_n; \beta_{n-2}|\beta_{n-1}),$$

(22)

where $I(\beta_n; \beta_{n-1})$ is the information on $\beta_n$ contained in $\beta_{n-1}$, and $I(\beta_n; \beta_{n-2}|\beta_{n-1})$ is the residual information on $\beta_n$ contained in $\beta_{n-2}$, once $\beta_{n-1}$ is known. A measure of the goodness of the one-step Markov approximation can be given in terms of

$$\zeta \geq \frac{I(\beta_n; \beta_{n-2}|\beta_{n-1})}{I(\beta_n; \beta_{n-1})}.$$  

(23)

In fact, if $\zeta \ll 1$ the relative importance of the numerator is small with respect to the denominator, meaning that, after $\beta_{n-1}$ is known, the additional information on $\beta_n$, carried by $\beta_{n-2}$, is negligible.

These quantities can be analytically computed from the above theory. The computations show that it is in fact $\zeta \ll 1$ for all values of the parameters we considered, proving that the packet success/failure process can be approximated very accurately by means of a Markov model, whose parameters can be readily found according to the above developments.

**A. Parameters of the Markov model**

Based on the above discussion, the process of the block successes and failures, $\beta_n$, can be modeled by means of a first-order Markov chain. Therefore, a Gilbert-Elliott channel at the bit level results in a simplified Gilbert channel [3] at the block level. This is an interesting result, since it allows us to take into account the physical layer channel description and its effect on the upper layers (data-link and network layers) by means of a very simple model, which lends itself to analytical developments.

Therefore, the channel is completely specified by the packet-level transition matrix with entries $P[\beta_n|\beta_{n-1}]$, i.e.,

$$P^{(x)} = \begin{pmatrix} P[0|0] & P[1|0] \\ P[0|1] & P[1|1] \end{pmatrix} = \begin{pmatrix} p^{(p)} & 1 - p^{(p)} \\ 1 - p^{(p)} & p^{(p)} \end{pmatrix},$$

(24)

with steady-state block error probability

$$\epsilon^{(p)} = P[\beta_1 = 1] = 1 - P[\beta_1 = 0] = \frac{1 - p^{(p)}}{1 - p^{(p)} + p^{(p)}},$$

(25)
In particular, we have that those at the symbol level can be computed from the above theory. Parameters of the Markov model at the packet level in terms of the previous block was in error, i.e., the steady-state probability of the channel being in a bad state is \( P_e(1) = 1 \) [3]. Similar results, not presented here due to lack of space, appear in [7] for the classic Gilbert channel, where \( P_e(1) = 0.5 \).

### A. Accuracy of the Markov model at the block level

In Fig. 1, we plotted the entropy of the binary process \( \beta_i, H(\beta) \) (solid), average mutual information, \( I(\beta_i; \beta_{i-1}) \) (dotted), and conditional average mutual information, \( I(\beta_i; \beta_{i-2}|\beta_{i-1}) \) (dashed), vs. the average burst length in symbols, \( 1/r \); average error rate \( \varepsilon = 0.01 \), no error correction, with \( * \) and without \( \circ \) interleaving, \( d = 10 \). \( N = 50, P_e(0) = 0, P_e(1) = 1 \).

### IV. RESULTS

In this section, we present some numerical results based on the theory presented above. For the sake of illustration, we consider the following values of the parameters: \( N = 50, P_e(0) = 0, P_e(1) = 1 \). This models the situation in which 50-symbol (e.g., bytes) packets are transmitted on a channel which follows the simplified Gilbert channel, where the symbol error probability in the bad state is \( P_e(1) = 1 \) [3]. Similar results, not presented here due to lack of space, appear in [7] for the classic Gilbert channel, where \( P_e(1) = 0.5 \).

#### 1. Accuracy of the Markov model at the block level

In Fig. 1, we plotted the entropy of the binary process \( \beta_i, H(\beta) \), the average mutual information, \( I(\beta_i; \beta_{i-1}) \), and the conditional average mutual information, \( I(\beta_i; \beta_{i-2}|\beta_{i-1}) \), vs. the average burst length in symbols, \( 1/r \); average error rate \( \varepsilon = 0.01 \) and no error correction capability \( (t = 0) \). It can be seen that in the presence of long enough bursts (say, more than twice the block length) the value of \( \zeta = I(\beta_i; \beta_{i-2}|\beta_{i-1})/I(\beta_i; \beta_{i-1}) \) is less than 1%, confirming the goodness of the Markov approximation. On the other hand, for bursts of moderate length the Markov approximation may become less accurate.

#### B. Parameters of the Markov model

From the above theory, it is possible to compute the probability of a block error, \( e(p) \), and the probability of a block success, given that the previous block was in error, \( r(p) \), regardless of the stochastic nature of the block error process. \( 1/r(p) \) has the meaning of average burst length. In all cases in which the error process at the packet level can be accurately modeled as Markov, these two quantities give its complete statistical description. However, even in the cases where the Markov model is not satisfactory, the above parameters are well defined and have physical significance, especially the average block error rate, \( e(p) \).

In Figs. 2 and 3, we plot \( e(p) \) and \( 1/r(p) \) vs. the average burst length (in symbols), for various values of the parameters. As the burst length increases the average block error rate tends to \( \pi_1 = \varepsilon \), i.e., the steady-state probability of the channel being in a bad state. In this case, when a symbol experiences bad (good) channel quality, it is very likely that all symbols in the block do, and the block is in error (correct) with high probability. As is evident from Fig. 3, in the presence of long bursts and without interleaving, the average burst length at the packet level is roughly equal to the burst length.
length in symbols divided by the number of symbols in a packet. This was also expected, since it is like counting the burst length in packets instead of symbols. On the other hand, in the presence of interleaving, the burst length at the packet level is practically equal to that at the bit level.

Fig. 2 also suggests that interleaving can be harmful. This is of course true in the absence of error correction. For example, if \( d < N \), a burst of \( d \) errors will result in \( d \) packets in error in the presence of interleaving, as opposed to at most \( 2 \) in its absence, and therefore the block error rate without interleaving is smaller. If some error correction is used, dispersing the errors might have a beneficial effect, if uncorrectable error patterns are mapped into correctable ones.

This can be clearly seen from Fig. 4, where we plotted the steady-state block error rate for the simplified Gilbert channel, for error correction capability \( t = 0 \) and \( 4 \), and for interleaving depth \( d = 1, 10, 100 \) and \( 1000 \). Intuitively speaking, in this case interleaving has the effect of reducing the burst length. In particular, for moderate values of the burst length (up to roughly \( d/10 \)), interleaving and error correction will make the channel look as if the errors were iid (which corresponds to the asymptotic behavior \( \varepsilon(p) \approx 1 - \sum_{i=0}^{t} \binom{N}{i} (1-p)^i p^{N-i} \approx \text{ steady-state} \)). As expected, Fig. 4 also shows that the advantage provided by interleaving is maintained only when \( td \) is larger than the average burst length; otherwise, interleaving is harmful, since it leads to uncorrectable error patterns in multiple packets.

C. Strategies for Error Control

In voice transmission, the low data rates encountered and the strict delivery time requirements justify the use of interleaving and of a moderate amount of FEC. On the other hand, when the data transmission rate is moderately high, the burst length in symbols may become very large, and an effective FEC and interleaving strategy will require interleavers with a very large depth and hence large memory. The delay introduced by interleavers and deinterleavers is about \( 2d \) packets. Hence the FEC and interleaving strategy, which is effective when \( td > 1/r \), will introduce a delay of the order of \( 1/rt \) packets.

In applications that can tolerate some jitter, such as data, image and to an extent video transfer, ARQ could be much more effective than the FEC and interleaving strategy. To begin with, in ARQ, the bandwidth overhead for error control is incurred only when needed, and is not wasted \textit{a priori} as in FEC. For example, consider the case shown in Fig. 4. For \( c = 0.001 \), it is not unusual to use a rate 1/2 FEC, to guarantee reasonable error rates after decoding, which results in a 50% throughput loss. On the other hand, for the block error rates of Fig. 4 (less than 0.01), the throughput decrease with an ARQ scheme is minimal. With regards to delay, ARQ will perform especially poorly if the propagation time is very large and the channel is error prone, since multiple transmissions of errored blocks lead to large overall delays. On the other hand, when the propagation time is small, as will be the case in micro/picocellular PCS systems, the delay involved with the use of an appropriately tuned ARQ will be of the order of the burst length \( 1/r \), i.e., not much larger with than FEC and interleaving (unless \( t \) is large, which on the other hand results in poor spectral efficiency). Furthermore, even if the application is sensitive to delay jitter, there is the possibility of using a smoothing buffer. These observations suggest that ARQ may well prove to be the error control mechanism of choice in the PCS environment [8].

V. CONCLUSIONS

In this paper we considered the residual bit/symbol error process and showed that the resulting block error process is well modeled as Markov. Parameters of the block error process were presented in terms of the parameters of the bit/symbol error process. These latter quantities can be determined based on the details of the physical environment. This work was motivated by the overall desire to determine how best to support various types of applications over bursty channels. We were able to characterize the range of data rates for which the use of FEC and interleavers helps.

The techniques described in this paper are useful in a number of other applications as well. An interesting extension would be to consider a number of possible channel states \( K > 2 \), with an error probability associated to each state. Such a model is believed to adequately model real-world channels, such as the fading radio channel [9]. The developments presented here can be readily modified to incorporate this feature.

REFERENCES