ARQ Error Control for Fading Mobile Radio Channels

Michele Zorzi, Member, IEEE, Ramesh R. Rao, Senior Member, IEEE, and Laurence B. Milstein, Fellow, IEEE

Abstract—In this paper, we study the correlation properties of the fading mobile radio channel. Based on these studies, we model the channel as a one-step Markov process whose transition probabilities are a function of the channel characteristics. Then we present the throughput performance of the Go-Back-N and selective-repeat automatic repeat request (ARQ) protocols with timer control, using the Markov model for both forward and feedback channels. This approximation is found to be very good, as confirmed by simulation results.

Index Terms—ARQ, correlated errors, fading channel, Markov model, time diversity, unreliable feedback.

I. INTRODUCTION

INTEREST in wireless communication has increased in recent years to foster personal communication services (PCS), as well as to explore the feasibility of ATM communications over wireless links. In this context, there is a need to integrate voice, data and other types of traffic over radio channels. This integration is likely to involve the segmentation of all types of traffic into blocks prior to their multiplexing on the link, and it is important that the system be robust enough to recover from the frequent errors that are likely to occur over a fading channel. Two factors will have a major impact on the performance of error control schemes over these channels, namely, the physical channel and the protocol. These two elements, usually studied separately at the physical and data-link layers, respectively, should be considered together for a successful deployment of effective error control strategies.

Therefore, the first need is to understand the effect of fading on the transmission of blocks of data. In particular, the modeling of the channel is a very important issue. In fact, the channel models used to compute the performance of the protocols at the data-link layer and above must reflect the physical layer characteristics in order for these results to be meaningful. Simplistic models or gross approximations (such as, for example, arbitrary independence assumptions) may lead to performance estimations which are far from reality. It may be even possible to take advantage of some features (e.g., the channel memory, which is usually destroyed by using interleaving). The availability of more detailed, but still tractable, channel models is therefore of primary importance.

Second, improved models for the performance assessment of error control schemes are to be considered. For example, most models that have been used for the study of automatic repeat request (ARQ) error control assumed that the block transmissions were independent and identically distributed (iid). In fact, protocols were specifically designed for iid channels, and techniques such as interleaving were developed to eliminate channel memory. In this context, approximating a channel with memory by means of a Markov model is more natural, and studying the performance of protocols on the Markov channel, and possibly developing new protocols suited for this kind of channel, is a more reasonable approach. In this paper, we will focus on ARQ error control protocols, which counteract errors through retransmission (as opposed to a priori protection such as forward error correction). In particular, we will consider Go-Back-N (GBN) and selective-repeat (SR) ARQ protocols with the use of timer control [1]–[6].

In most of the existing literature on ARQ, errors in distinct blocks are assumed independent, and feedback is assumed to be error-free. These conditions are not satisfied in a mobile radio environment. However, some papers that discuss the performance of ARQ protocols under more general conditions have appeared. Kanal and Sastry [7] review the dependent channel, but focus more on forward error correction (FEC) techniques. Towsley [8] considers a Markov forward channel model, and focuses on queueing performance. Leung et al. [9] also give results for the ideal feedback case. A generalized Markov model of higher order is considered in [10], but the effect of unreliable feedback is not included. Kim and Un appear to have been the first to study the effect of a Markov model for the feedback channel as well [11], combining the effects of dependent transmissions and unreliable feedback. Similar results are found for some more complex protocols, which include control mechanisms to resolve uncertainties that may arise when feedback information is lost [4]. In this last paper, for analytical convenience, some rules of the actual protocols are disregarded in the analysis, leading to an upper bound for the throughput [5]. A simulation study of GBN and SR in Rayleigh fading channels has also been presented in [2] and [3], along with a throughput estimation technique. Finally, in [5], [12], and [13], an analytical technique based on a Markov chain analysis is presented, which allows one to accurately study the performance of the GBN protocol with timer control in Markov channels with unreliable feedback. This technique has been employed to study SR as well [6], and can be potentially extended to the study of other ARQ protocols involving memory [14].

Manuscript received July 12, 1995; revised March 11, 1996 and June 23, 1996. Parts of this work have been presented at ICU'95, Tokyo, Japan, Nov. 1995, and at MILCOM'95, San Diego, CA, Nov. 1995.

The authors are with the Center for Wireless Communications and the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92037-0407 (e-mail: zorzi@ece.ucsd.edu, rao@ece.ucsd.edu, milstein@ece.ucsd.edu).

Publisher Item Identifier S 0018-9545(97)03124-1.
In this paper, we study the correlation properties of the fading process, as a prelude to analyzing the performance of ARQ protocols. Recently, the Markov character of the envelope of a Rayleigh fading process has been assessed [15]. Similar results are discussed in this paper, with reference to the binary process of successes/failures on the channel. It is shown that in most cases of practical interest, this process can be modeled as a binary Markov channel (similar results for packet transmission on a Gilbert channel [16], commonly used to model symbol-error bursts, are illustrated in [17]). As a second contribution, we apply the above channel model and the analysis in [5], [6], and [13] (which applies to general Markov channels with unreliable feedback) to study the throughput performance of GBN and SR with timer control over fading channels, and to investigate the effect of the system parameters.

Our approach, which spans both the physical layer (the channel) and the data-link/network layers (the protocol), is organized as follows. In Section II, the basic ARQ protocols we consider in this paper are briefly described. In Section III, the fading model is presented. The process of successes and failures of data block transmissions is introduced in Section IV, and a Markov approximation of it is proposed, discussed and validated. Performance results for ARQ protocols are given in Section V. Finally, Section VI concludes the paper.

II. BASIC ARQ PROTOCOLS

Three basic protocols are commonly considered [1]: stop-and-wait (SW), Go-Back-N, and selective-repeat. SW is less efficient than the other two, and cannot take advantage of the time-diversity feature. Therefore, in the following we will only consider GBN and SR with time diversity. For conciseness, we only give a brief outline of the protocol rules. The interested reader can find more details in [1]–[6] and [18].

Packet lengths are assumed to be a constant, equal to one time unit (slot). The round-trip delay is \( m + 1 \) slots from the end of a transmission to the reception and decoding of the corresponding feedback information, so that if a packet transmitted in slot \( i \) is negatively acknowledged, it will be retransmitted in slot \( i + m \). It is assumed that positive (ACK) and negative (NAK) acknowledgment can never be confused with each other, i.e., the effect of backward errors is to map the ACK and NAK symbols to an erasure symbol. Also, each ACK/NAK carries the identity of the last correctly received packets. This implies that a packet whose ACK is lost may be subsequently acknowledged by feedback received in the future.

A. Go-Back-N

Consider the Go-Back-N ARQ protocol with timer control, as described in [4]. The receiver follows the standard rules of ARQ GBN [18], i.e., it sends an ACK for every correctly received packet. When it receives an incorrect packet, it sends a NAK instead, and discards every successive packet, until a correct copy of the negatively acknowledged packet is received. The transmitter acts according to the following rules:

1. It sends packets in order, as long as it receives ACK's on the backward channel.
2. Upon reception of NAK, it goes back to packet \( i \), retransmitting in order all packets from packet \( i \).
3. Garbled feedback, when received, is ignored. A future ACK/NAK may provide information that could acknowledge packet \( i \). In fact, the ACK/NAK of packet \( k \) also contains implicit acknowledgment of all packets \( i < k \). However, it is possible that the lost feedback was a NAK (in which case no more feedback will be sent, and the uncertainty could last for ever), or that the feedback channel is undergoing a very long burst of errors, so that a large number of ACK's get lost, and the resulting delay is unacceptable. To avoid such cases, a time-out mechanism is used. This causes the transmitter to retransmit packet \( i \) \( t \) slots later (and all packets following \( i \) afterwards), if by that time it is still unknown whether or not packet \( i \) was correctly received. This allows the transmitter to avoid deadlock situations or buffer overflow (in particular, the transmitter buffer needs only contain \( t \) packets). Of course, it must be \( t \geq m \) (if \( t = m \) we have the classic GBN scheme, in which if an ACK is not received at the proper time, retransmission is immediately performed).

B. Selective-Repeat

We consider a generalization of the standard version of the SR ARQ strategy [19]. The transmitter transmits blocks, whose receptions are acknowledged by the receiver, by means of a feedback message that can be an ACK or an NAK. Blocks are kept in a buffer at the transmitter, and removed from it when acknowledged. NAK’s trigger retransmission of the corresponding blocks. It is assumed that, under correct operating conditions, the feedback is known at the transmitter \( m \) slots (round-trip delay) after the block transmission started. In the classic version of SR, the absence of feedback is equivalent to an NAK, and the block is immediately retransmitted. In the presence of feedback errors, however, waiting for future feedback information, rather than immediately retransmitting, may allow the recovery of some lost acknowledgment, thereby improving the performance.

In order to exploit this, a time-out counter is assigned to each block, and activated when the block is transmitted. When the counter reaches \( t \) slots, the time-out expires, and a retransmission is scheduled in the next slot. The classic version of SR corresponds to \( t = m \); whenever \( t > m \), some time-diversity is introduced, allowing feedback error recovery. In particular, each correct feedback message is assumed to contain ACK’s and NAK’s about all past blocks (for example, it could identify all erroneous outstanding blocks). Upon the correct reception of such a feedback message, the erroneous blocks are rescheduled for retransmission, and the correct ones are removed from the transmitter memory.

This mode of operation requires some buffer capacity at the receiver. Once a block is in error, all successive blocks must be retained in memory, until the original block is correctly received. This is to ensure that blocks are released in order
to the upper layers. Therefore, the ordering of packets is achieved at the price of large memory and possibly long delays. To mitigate this problem, some modified schemes have been proposed [19]–[21]. Here, we are interested in the basic protocol, and therefore we assume an infinite buffer length.

### III. Fading Model

In this paper, we assume that the data-rate is relatively high, so that the duration of a bit (and even of a packet) is smaller than the coherence time of the channel (i.e., the time scale of the fading variations). We also assume a flat fading channel, modeling it as a multiplicative complex function, \( \alpha(t) \), which is adequately described as a random process. A popular model considers a Gaussian random process with a given mean and covariance function [22]. On the time scale of the fading variations, the process can and will be considered as stationary. Therefore, with no loss in generality, we will normalize its power to one. The real and imaginary axes can be chosen so that the mean \( \mu = E[\alpha(t)] \) is real. Also, we consider the covariance function, defined as

\[
K(\tau) = E[(\alpha(t) - \mu)(\alpha(t + \tau) - \mu)].
\]

Note that if \( \mu = 0 \), the envelope of \( \alpha(t) \) is Rayleigh distributed for any \( t \), and the envelope squared has an exponential distribution. On the other hand, when \( \mu > 0 \), we are in the presence of Rician fading: this latter model accounts for the presence of a line-of-sight (LOS) component, and is often more accurate in micro- and picocells. When the LOS component is absent, or has negligible power, the Rician model degenerates into the Rayleigh one.

In a widely accepted model, the Gaussian process is assumed to have a bandlimited nonrational spectrum, given by [22]

\[
S(f) = S(0) \left[ 1 - \left( \frac{f}{f_D} \right)^2 \right]^{-1/2}, \quad \text{for } |f| < f_D \tag{2}
\]

and zero, otherwise. The parameter \( f_D \) is called the Doppler frequency, and is inversely proportional to the coherence time. The spectrum (2) corresponds to the covariance function

\[
K(\tau) = J_0(2\pi f_D |\tau|) \tag{3}
\]

whose physical meaning has been investigated in [22] and [23]. \( J_0(\cdot) \) is the modified Bessel function of the first kind and of zeroth order. Note that (3) corresponds to a process with unit power, i.e., \( E[|\alpha(t)|^2] = 1 \). This implies that

\[
S(0) = \frac{2}{\pi} \tag{4}
\]

Note also that the mean and the covariance, along with the assumption that the real and imaginary parts are independent, give a complete statistical description of the process, since it is Gaussian.

The covariance function, \( K(\tau) \), is plotted versus \( f_D |\tau| \) in Fig. 1. The correlation properties of the fading process depend in fact only on this parameter. When \( f_D |\tau| \) is small (\(< 0.1\)) the process is very correlated ("slow" fading); on the other hand, for larger values of \( f_D |\tau| (> 0.2) \), two samples of the channel are almost independent ("fast" fading).

Note that, for high data rates (small \( \tau \), the fading process can always be considered to be slowly varying, at least for the usual values of the carrier frequency (900–1800 MHz) and nominal vehicular speeds. Therefore, the dependence between transmissions of consecutive blocks of data cannot be neglected. In particular, the assumption that the successes/failures of data blocks constitute an iid process is far from reality, and may lead to incorrect results when used to evaluate the performance of a transmission scheme or protocol. A more general model for the success/failure process, which accounts for dependence, is developed in the next section.

### IV. A Model for the Success/Failure Process

Let us now consider sampling the fading process. This is natural when considering block or packet transmission. As an example, consider a TDMA system, in which packets are transmitted in frames. The channel can be considered constant during a packet (especially if the packet is short and the data rate is sufficiently high), whereas it is not so at points separated by a frame distance (i.e., for two consecutive packets on the same channel). In this case, the sequence of the channel values experienced by the packets on a TDMA channel can be seen as a sampled version of the fading process at the frame frequency. As another example, consider a continuous stream of packets on a channel, with the packet transmission rate much larger than the Doppler frequency. In this case, it is reasonable to assume that the various bits of a same packet, and even two consecutive packets, will experience approximately the same channel conditions. On the other hand, the channel is not exactly constant, and therefore a dynamic model must be developed to capture its long-term evolution.

According to the above model, a vector of \( n \) channel samples

\[
\Omega_n = (\alpha_1, \alpha_2, \cdots, \alpha_n), \quad \alpha_i = \alpha(iT) \tag{5}
\]

is a complex Gaussian random vector with known mean and covariance matrix, \( \mathbf{K} \). The joint probability density function (pdf) is known, and the conditional pdf of the most recent
sample, $\alpha_n$, given all the others, is

$$P(\alpha_n|\alpha_{n-1}) = \frac{1}{\sqrt{2\pi}\sigma_n^2(\alpha_{n-1})} \exp\left[-\frac{|\alpha_n - \mu_n(\alpha_{n-1})|^2}{2\sigma_n^2(\alpha_{n-1})}\right]$$

(6)

where the dependence of the mean and variance on the past samples is highlighted. The functional dependence of $\mu_n$ and $\sigma_n$ on $\alpha_i$, $i = 1, \ldots, n-1$, can be found by using standard linear prediction theory [24].

Recently, it has been shown by Wang [15] that a one-step Markov model adequately describes the discrete process $v_i = |\alpha_i|$. On the other hand, from a communications point of view, the relevant quantity is not actually the channel amplitude, $v(t)$, but rather some function of it; for example, the error probability of a block of bits, which is a nonlinear function of $v(t)$, according to the modulation format and the coding scheme [24]. In this case, we can define a new random process $\beta(t)$, which depends on $v(t)$, as

$$\beta(t) = \varphi[v(t)]$$

(7)

in which, for simplicity, we assume an instantaneous (i.e., memoryless) relationship between the two processes.

As an example, consider the following approximation [2], whose accuracy is discussed in Section IV-C below. In this scheme, the success of a block is determined by comparing the signal power to a threshold: if the received power is above a certain threshold, the block is successfully decoded with probability one; otherwise, it is lost with probability one. In this case, $\varphi$ is a step function, and the binary process which describes block successes and failures on the channel, $\beta_i$, is obtained by quantization of the magnitude squared of the complex Gaussian description, i.e.,

$$\beta_i = \begin{cases} 0, & \text{if } |\alpha_i|^2 > b \\ 1, & \text{if } |\alpha_i|^2 \leq b \end{cases}$$

(8)

where $b$ is the power threshold, and “1” stands for a packet failure. The quantity $F = 1/b$ is called the fading margin, and is the maximum fading attenuation that the system can tolerate.

Note that, even if the sequence $v_i$ were modeled as a Markov process (which has been shown to be accurate [15]), this would not be necessarily true for the new sequence, $\beta_i$, since a quantized version of a Markov process may not be Markov. However, the possibility of predicting $\beta_i$ based only on a limited (and possibly small) number of previous samples deserves more investigation: it can be conjectured, in fact, that, for highly correlated fading, the quantized version will also be highly correlated, allowing for a simplified model. This possibility is also appealing because there exist in the literature many results which apply to Markov channels, and which could be used in the present context. In the following, we explore the Markov approximation, and show that it turns out to adequately describe the channel.

Following Wang [15], we use mutual information to show that the success/failure of the transmission in the previous slot summarizes almost all the information contained in the past. This approach was used by Wang with reference to the values of the fading envelope, $v_i = |\alpha_i|$. Here, we will apply the same technique to the success–failure process $\beta_i$.

Let $I(\beta_i; \beta_{i-1}; \beta_{i-2})$ be the average mutual information between the random variable $\beta_i$ and the past two transmissions $\beta_{i-1}$ and $\beta_{i-2}$. We can write [15]

$$I(\beta_i; \beta_{i-1}; \beta_{i-2}) = I(\beta_i; \beta_{i-1}) + I(\beta_i; \beta_{i-2} | \beta_{i-1})$$

(9)

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

$$\zeta = \frac{I(\beta_i; \beta_{i-2} | \beta_{i-1})}{I(\beta_i; \beta_{i-1})}.$$  

(10)

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

In [15], the joint pdf of three successive envelope samples is found to be

$$f_{v_{i-2}v_{i-1}v_i}(a_1, a_2, a_3) = \frac{\rho a_1 a_2 a_3}{\det[A]} \exp\left[-\frac{1}{2}(\eta_{11} a_1^2 + \eta_{22} a_2^2 + \eta_{33} a_3^2)ight]$$

$$\cdot \sum_{\Omega_{ji,k}} (-\eta_{12} a_1 a_2) (-\eta_{23} a_2 a_3) (-\eta_{31} a_3 a_1)^k$$

$$\cdot \left[\sum_{l=\min}^{l=\max} \left(\frac{j}{l} - \frac{i}{l} - \frac{k}{l}\right) \left(\frac{j}{l} - \frac{i}{l} - \frac{k}{l}\right) \right]$$

(11)

where $\Omega_{ji,k}$ is the set of all triples of nonnegative integers $i$, $j$, and $k$, which are either all even or all odd

$$l_{\min} = \max\left(0, \frac{i-j}{2}, \frac{i-k}{2}\right)$$

(12)

and

$$l_{\max} = \min\left(i, \frac{i+j}{2}, \frac{i+k}{2}\right).$$

(13)

Also

$$A = \frac{1}{2} \begin{bmatrix} 1 & J_1 & J_2 \\ J_1 & 1 & J_1 \\ J_2 & J_1 & 1 \end{bmatrix}$$

(14)

where $J_1 = J_0(2\pi f_P T)$ and $J_2 = J_0(4\pi f_P T)$ (the factor 1/2 is the power of each orthogonal Gaussian component, so that the power of the envelope is equal to one, and the $\eta_{ji,k}$'s are defined so that

$$A^{-1} = \begin{bmatrix} \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{31} & \eta_{32} & \eta_{33} \end{bmatrix}.$$  

(15)

With a slight change of notation, (11) can be rewritten as

$$f_{v_{i-2}v_{i-1}v_i}(a_1, a_2, a_3) = \sum_{\Omega_{ji,k},l=0}^{\max} [\xi_{\Omega_{j+1,k+1}}(a_1, \eta_{11})$$

$$\psi_{j+k}(a_2, \eta_{22})]_{\psi_{j+k+1}(a_3, \eta_{33})}$$

$$- \xi_{\Omega_{j+1,k+1}}(a_1, \eta_{11})$$

$$\psi_{j+k}(a_2, \eta_{22})]_{\psi_{j+k+1}(a_3, \eta_{33})}$$

$$+ \sum_{\Omega_{ji,k},l=0}^{\max} [\xi_{\Omega_{j+1,k}}(a_1, \eta_{11})$$

$$\psi_{j+k}(a_2, \eta_{22})]_{\psi_{j+k+1}(a_3, \eta_{33})}$$

$$- \xi_{\Omega_{j+1,k}}(a_1, \eta_{11})$$

$$\psi_{j+k}(a_2, \eta_{22})]_{\psi_{j+k+1}(a_3, \eta_{33})}$$

(16)
where
\[ \xi_{even}(i, j, k) = \frac{\eta_{i}^{2} \eta_{j}^{2} \eta_{k}^{2}}{(2i+1)(2j+1)(2k+1) \det[A]} \sum_{l= \min}^{\max} \binom{2i}{l} \binom{2j}{l-i} \binom{2k}{l-i+j} \] (17)

and
\[ \xi_{odd}(i, j, k) = \frac{\eta_{i}^{2} \eta_{j}^{2} \eta_{k}^{2}}{(2i+1)(2j+1)(2k+1) \det[A]} \sum_{l= \min}^{\max} \binom{2i+1}{l} \binom{2j+1}{l-i} \binom{2k+1}{l-i+j} \] (18)

With
\[ f_{\text{min}}^{(e)} = \min(i, j, k), \]
\[ f_{\text{max}}^{(e)} = \max(i, j, k) + 1, \]
\[ f_{\text{min}}^{(o)} = \max(0, i - j, i - k), \]
\[ f_{\text{max}}^{(o)} = \min(2i, i + j, i + k), \]

and
\[ \psi_n(a, \eta) = a(\frac{\alpha^2}{2})^n e^{-\eta \alpha^2/2}. \] (21)

With \( \beta_i = 0 \) for \( \eta_i^2 > \beta \) and one otherwise, the joint probability mass function of \( \beta_1/\beta_2/\beta_3 \) can be found as
\[ P[\beta_1 = \gamma_1, \beta_2 = \gamma_2, \beta_3 = \gamma_3] = \int_{\Gamma_1} d\alpha_1 \int_{\Gamma_2} d\alpha_2 \int_{\Gamma_3} d\alpha_3 \sum_{k=0}^{\infty} \psi_{\gamma_1}(\alpha_1, \alpha_2, \alpha_3) \]
\[ = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left[ \psi_{\gamma_1}(i, j, k) \psi_{\gamma_1+k}(i, \eta_{11}) \right. \]
\[ \cdot \int_{\Gamma_1} d\alpha_1 \psi_{i+j+k}(\alpha_1, \eta_{12}) \int_{\Gamma_2} d\alpha_2 \psi_{j+k}(\alpha_2, \eta_{23}) \]
\[ - \xi_{odd}(i, j, k) \int_{\Gamma_1} d\alpha_1 \psi_{i+j+k+1}(\alpha_1, \eta_{11}) \]
\[ \left. \cdot \int_{\Gamma_2} d\alpha_2 \psi_{i+j+k+1}(\alpha_2, \eta_{22}) \int_{\Gamma_3} d\alpha_3 \psi_{j+k+1}(\alpha_3, \eta_{33}) \right] \] (22)

where
\[ \Gamma_i = \begin{cases} [0, \sqrt{\beta}], & \text{for } \gamma_i = 1 \\ (\sqrt{\beta}, \infty), & \text{for } \gamma_i = 0. \end{cases} \] (23)

It is easy to see that the integrals in the summation can be analytically solved as
\[ \Phi_n(\beta_i, \eta) = \int_{0}^{\infty} \psi_n(a, \eta) da \]
\[ = \int_{0}^{\infty} a^n e^{-\eta a^2} da \]
\[ = \frac{\eta^n k!}{2^n} \sum_{k=0}^{n} \frac{\eta^n k!}{k!} \frac{(\eta/2)^k}{(k+1)^n} \] (24)

Finally, the joint probability mass in (22) can be expressed as follows:
\[ P[\beta_1 = \gamma_1, \beta_2 = \gamma_2, \beta_3 = \gamma_3] \]
\[ = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left[ \psi_{\gamma_1}(i, j, k) \psi_{\gamma_1+k}(i, \eta_{11}) \right. \]
\[ \cdot \int_{\Gamma_1} d\alpha_1 \psi_{i+j+k}(\alpha_1, \eta_{12}) \int_{\Gamma_2} d\alpha_2 \psi_{j+k}(\alpha_2, \eta_{23}) \]
\[ - \xi_{odd}(i, j, k) \int_{\Gamma_1} d\alpha_1 \psi_{i+j+k+1}(\alpha_1, \eta_{11}) \]
\[ \left. \cdot \int_{\Gamma_2} d\alpha_2 \psi_{i+j+k+1}(\alpha_2, \eta_{22}) \int_{\Gamma_3} d\alpha_3 \psi_{j+k+1}(\alpha_3, \eta_{33}) \right] \] (25)

where \( \Xi^{(i)} = \Xi \) for \( \eta_i = 1 \) and \( \Xi^{(i)} = \Phi \) for \( \eta_i = 0 \).

Numerical results have been obtained both from the above Gaussian approach and through direct simulation of the fading process, as in [22]. For example, Fig. 2 shows \( I(\beta_i; \beta_{i-1}) \) and \( I(\beta_i; \beta_{i-2}; \beta_{i-1}) \) versus the fading margin, \( F \) (in dB); the entropy of the source, \( H(\beta) \), is also shown. Two cases are considered here: slow fading \( (f_{D}T = 0.02144) \) and fast fading \( (f_{D}T = 1.28) \) Rayleigh fading. Curves correspond to the analytical Gaussian model, symbols to simulation results [22].
in approximating the data block success/failure process on a fading mobile radio channel.

Therefore, in what follows, the pattern of packet errors is assumed to follow a first-order Markov model, described by the transition matrix \( M(x) = M(1)^x \), with

\[
M(x) = \begin{bmatrix}
    p(x) & q(x) \\
    r(x) & s(x)
\end{bmatrix} \\
M(1) = \begin{bmatrix}
    p & q \\
    r & s
\end{bmatrix}
\] (27)

where \( p(x) = 1 - q(x) \) and \( r = 1 - s(x) \) are the probabilities that the transmission in slot \( x \) is successful, given that the transmission in slot \( i \rightarrow x \) was successful (unsuccessful). Note that \( 1/r \) represents the average length of a burst of errors, which is described by a geometric random variable.

Given the matrix \( M(1) \), the channel properties are completely characterized. In particular, it is possible to find the marginal probability that a packet is in error, \( \varepsilon \), given by

\[
\varepsilon = 1 - \frac{r}{1 - p + r}. \tag{28}
\]

A similar model applies to the feedback channel.

A. Computation of the Parameters

In order to carry out the performance analysis, the parameters of the Markov model presented in the previous section must be found. From the fading model presented above, it is fairly easy to write an analytical expression for the transition probabilities of a Markov model. In general, the computation of the transition probabilities of an \( n \)-th order Markov model involves nonlinear functions of 2\( n \) Gaussian variables, some of which are dependent upon each other. A Monte Carlo evaluation of the probabilities may be a more convenient solution, even though it is still practically limited by the achievable accuracy, since some joint probabilities can be very small. Alternatively, a fading generator can be used to simulate the channel and to directly measure the parameters of interest. In this section, we discuss both approaches to this problem, analysis, and simulation.

In the text, we will refer to the forward channel; of course, everything is the same with reference to the backward link. There are essentially two independent parameters to be computed, e.g., \( p \) and \( r \). Following most of the literature, we will study \( \varepsilon \) and \( r \), since \( p \) can be found from (28). This choice is motivated by the fact that these two parameters have an immediate physical significance: \( \varepsilon \), as already mentioned, is the average block error rate, and measures how often a block in error, \( 1/r \) is the average length of an error burst, and gives an idea about how clustered the errors tend to be.

1) Analytical Approach: Let \( \mathcal{A}(\nu) \) be the event of a successful reception, conditioned on the value of the fading envelope, \( \nu = [\alpha] \), so that the conditional probability that a block is received in error (WER) is given by

\[
P_{w}(\nu) = 1 - P[\mathcal{A}(\nu)]. \tag{29}
\]

The average probability of a block error is therefore

\[
\varepsilon = \mathbb{E}[P_{w}(v)] = \int_{0}^{\infty} P_{w}(\alpha) f_{w}(\alpha) \, d\alpha.
\] (30)

where the expectation is performed with respect to the pdf of the fading envelope, \( f_{w}(\alpha) \).

The probability that two successive blocks are in error is given by

\[
P[1,1] = \mathbb{E}[P_{w}(\nu_1)P_{w}(\nu_2)] = \int_{0}^{\infty} \int_{0}^{\infty} P_{w}(\alpha_1)P_{w}(\alpha_2)f_{\nu_1\nu_2}(\alpha_1, \alpha_2) \, d\alpha_1 \, d\alpha_2
\]

(31)

and

\[
r \equiv 1 - P[1,1] = 1 - \frac{P[1,1]}{\varepsilon}.
\] (32)

If second-order selection combining diversity is used, whereby the fading values at the two antennas are independent of each other, then we have for the conditional probability of unsuccessful reception

\[
P_{d}(\nu) = 1 - P[\mathcal{A}(\nu)]
\]

with \( \nu = \max \{\nu^{(1)}, \nu^{(2)}\} \). If \( F_{\nu}(a) = P[\nu^{(1)} \leq a] = P[\nu^{(2)} \leq a] \) is the common cumulative density function (cdf) of the fading envelope at each antenna, we have

\[
F_{\nu}(a) = [F_{\nu}(a)]^2
\]

(34)

and

\[
\varepsilon_{d} = \mathbb{E}[P_{d}(\nu)] = \int_{0}^{\infty} P_{d}(\alpha)2F_{\nu}(a)f_{\nu}(\alpha) \, d\alpha
\]

(35)

where the subscript “\( d \)” denotes the quantities for the diversity case. Also,

\[
F_{\nu_{1}\nu_{2}}(a_1, a_2) = [F_{\nu_{1}\nu_{2}}(a_1, a_2)]^2
\]

(36)

and

\[
P_{d}(1,1) = \mathbb{E}[P_{w}(\nu_1)P_{w}(\nu_2)] = \int_{0}^{\infty} \int_{0}^{\infty} P_{w}(\alpha_1)P_{w}(\alpha_2)f_{\nu_{1}\nu_{2}}(a_1, a_2) \, d\alpha_1 \, d\alpha_2.
\]

(37)

If

\[
P_{w}(\nu) = \begin{cases} 0, & \nu^2 > b \\ 1, & \nu^2 \leq b \end{cases}
\]

(38)

we have

\[
\varepsilon = \varepsilon_{d} \tag{39}
\]

\[
P[1,1] = P_{\nu_{1}\nu_{2}}(\sqrt{b}, \sqrt{b}) \tag{40}
\]

\[
\varepsilon_{d} = \varepsilon^2 \tag{41}
\]

\[
P_{d}[1,1] = [F_{\nu_{1}\nu_{2}}(\sqrt{b}, \sqrt{b})]^2 = (P[1,1])^2 \tag{42}
\]
For the case of Rayleigh fading, the pdf of the envelope is
\[ f_x(x) = 2 \alpha e^{-\alpha^2} \]
and the joint density is
\[ f_{x_1 x_2}(x_1, x_2) = \frac{\alpha_1 \alpha_2}{1 - \rho^2} e^{-(\alpha_1^2 + \alpha_2^2)/2(1 - \rho^2)} I_0 \left( \frac{\rho \alpha_1 \alpha_2}{1 - \rho^2} \right) \]
where \( \rho = J_0(2\pi f_D T) \) is the correlation coefficient of the two successive samples, \( T \) is the sampling time, \( f_D \) is the Doppler frequency, and \( J_0 \) and \( f_0 \) are the Bessel function and the modified Bessel function of the first kind, respectively, both of zeroth order. Similar distributions can be found for the case of Rician fading [25].

For the Rayleigh fading case, \( \varepsilon \) and \( r \) can be shown to be given by
\[ \varepsilon = 1 - e^{-b} \]
and
\[ r = \frac{Q(\theta, \rho \theta) - Q(\rho \theta, \theta)}{e^b - 1} \]
respectively, where
\[ \theta = \sqrt{\frac{2b}{1 - \rho^2}} \]
and \( Q(\cdot, \cdot) \) is the Marcum \( Q \) function.

2) Simulation Approach: It is also possible to directly simulate the fading process, using some of the techniques suggested in [22] and [26]. For example, the simulation results we obtained in the following were computed by sampling the fading process and checking the threshold condition (8) to obtain the error process. The transition probabilities of this process can then be evaluated by considering a long run of this sequence. A more detailed approach would be to consider the actual word error rate, instead of the threshold model. As discussed in Section IV-C below, however, this more complex procedure leads to results which do not differ substantially from those obtained with the threshold model, confirming the wide applicability of the threshold model itself.

The simulation approach (as opposed to the analytical one above) is valuable in at least two ways: 1) it provides a way to quickly compute the probabilities, which in some cases might even be faster than numerical integration and 2) it applies to virtually any environment and to any model of fading, where the fading envelope distribution can even be unknown. Note that it is also possible to obtain the Markov parameters by direct field measurements.

B. Example of Application

As an example, we computed the performance of the GBN protocol with perfect feedback, for the two cases considered above: slow fading \( (f_D T = 0.02144) \), and fast fading \( (f_D T = 1.28) \) [2], [3]. The success/failure of a block is assumed to be found as in (8). The GBN throughput, for a round-trip delay of \( m - 1 = 4 \) slots, is shown in Fig. 3 versus the value of the fading margin \( F \), for several values of the memory, \( n \), and under the assumption of slow fading. Also marked in Fig. 3 are some simulation points. It can be seen that in the absence of memory (i.e., assuming iid slot outcomes), the performance is significantly underestimated with respect to the models which take into account some channel history. On the other hand, it can be seen that considering further slots besides the latest one adds very little to the throughput estimation, indicating that, as far as our performance evaluation is concerned, the channel can be satisfactorily modeled as one-step Markov. This is even more true in the case of fast fading, in which consecutive slots are almost independent. From Table I, which shows the throughput performance of the GBN protocol versus the fading margin and the order of the Markov model, \( n \), it can be seen that, even in the absence of memory, the throughput estimation yields almost the same result. As expected, this shows that the iid approximation is very good.

C. Accuracy of the Threshold Model

In the above analysis and simulation models, we considered the threshold model (8) to determine the success or failure of a packet transmission. Given the quantity \( f_D T \), i.e., the

![Fig. 3. Throughput of GBN with perfect feedback η versus the fading margin F, for some value of the memory of the Markov model n (markers = simulation), m = 5, and slow fading (f_D T = 0.2144).](image1)

<table>
<thead>
<tr>
<th>memory</th>
<th>fading margin, F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 dB</td>
</tr>
<tr>
<td>0</td>
<td>0.539</td>
</tr>
<tr>
<td>1</td>
<td>0.540</td>
</tr>
<tr>
<td>2</td>
<td>0.540</td>
</tr>
<tr>
<td>3</td>
<td>0.546</td>
</tr>
<tr>
<td>simulation</td>
<td>0.546</td>
</tr>
</tbody>
</table>

| Table I | Throughput of GBN with Perfect Feedback: Markov Model with Different Memory and Simulation Results Compared; m = 5, F = 8, 13, and 18 dB, Fast Fading (f_D T = 1.28) |
fading variability, this approach allowed us to summarize the channel conditions by means of a single parameter, i.e., the fading margin, $F$, or (equivalently) the threshold, $1/F$. Also, the analysis is much simpler in this case than it would be if we took into account the details of the modulation/coding technique. In this subsection, we discuss the accuracy of this simpler model.

Consider a given value of the channel variability, $f_D T$. Note that the Markov model has two independent parameters, $\varepsilon$ and $r$, so that, in general, it is not true that a single quantity (the fading margin, in this case) is sufficient to characterize the channel. Consider, for example, a general coding/modulation scheme and a threshold model. For any value of the average signal-to-noise ratio of theformer, it is possible to pick a value of the threshold of the latter so that they result in the same packet error rate, $\varepsilon_r$. However, this does not guarantee, in general, that the value of $r$ will be the same in the two cases.

To further investigate the issue, let us focus on slow fading first, where the fading value can be assumed constant throughout a packet. We ran a simulation where the fading process (simulated according to Jakes [22]) was sampled to give the value of the fading envelope, $\psi$, in each slot. We assumed a packet length of $n$ bits, a block code having an error correction capability of $c$ bits per packet, and a bit error rate $P_e(\psi)$, conditioned on the instantaneous value of the fading envelope, $\psi$, so that the conditional packet error probability, $P_{\psi}(\psi)$, is given by

$$P_{\psi}(\psi) = \sum_{i=0}^{n} \binom{n}{i} P_e(\psi)^i (1 - P_e(\psi))^{n-i}. \quad (49)$$

From this simulation, for different values of the parameters, we could determine the second-order statistics of the error process, and, in particular, the values of $\varepsilon$ and $r$. If we use a step function for $P_{\psi}(\psi)$, instead of (49), we find the analogous results for the threshold model. We considered the relationship between $r$ versus $\varepsilon$ (obtained by varying the average signal-to-noise ratio), for various values of $n$ and $c$ and for various modulation schemes [which result in different expressions for $P_e(\psi)$], and found that its dependence on those parameters is not strong. This means that, given the value of the average block error rate, $\varepsilon$, the second-order statistics is almost independent of the coding/modulation scheme used, and is very close to that found by using the threshold model. Moreover, our results showed that the performance of the ARQ schemes exhibits rather weak dependence on $r$, so that the differences in throughput are even smaller. As an example, the insensitivity of the throughput performance to the error correction capability of the code is shown in Fig. 4, for 500-b noncoherent FSK modulated packets (where $P_e = 0.2 e^{-\Delta SNR/2}$), for various values of $c$ and for the threshold model. It is clear that the results obtained for the threshold model, very simple to compute, can be used as "universal curves," so that lengthy simulations can be avoided. Entirely similar results (not shown here) have been obtained in for coherent BPSK and for noncoherent $M$-ary FSK.

An analogous (although more complicated) approach could be developed for fast fading. However, if the fading is sufficiently fast, the packet errors tend quickly to be iid, so that $r \approx 1 - \varepsilon$ for all schemes. More details and results have been reported in [27].

V. NUMERICAL RESULTS

In this section, some numerical results are presented for the Markov parameters of the channel, and for the performance of both GBN and SR, under a number of different conditions. These results are obtained using the analysis in [5], [6], and [13] for the throughput performance of GBN and SR with timer control. These results apply to a general one-step Markov model for both forward and feedback channels. Thus, they can be used in the present context with the specific transition probabilities derived in Section IV based on the threshold model. All results to be presented are obtained through Markov analysis. The actual fading process was also simulated (as proposed in [22]), and was sampled to obtain the value of the fading envelope for each data block which was compared to the threshold to obtain the error process. These simulation results (not shown in the graphs to improve readability) were always in good agreement with the analytical ones. Also, simulated throughput results were found to yield a slightly better performance than analysis.

In Fig. 5, the Markov parameters of the channel, $\varepsilon$, $r$, and $1 - p$ are plotted versus the fading margin $F$. It can be seen the beneficial effect of the use of diversity; $\varepsilon$ is reduced (less errors), $r$ is increased (shorter error bursts), and $1 - p$ is smaller (longer runs of error-free transmissions). Also, comparing Fig. 5(a) (slow fading) and (b) (fast fading), it is clear that $r$ and $1 - p$ are larger for fast fading than for slow fading. In particular, the following conclusions can be drawn: $\varepsilon$ is the same in the two cases (it is determined by the first-order statistics only); $1 - p$ (conditional probability of error) for fast fading is practically equal to $\varepsilon$ (unconditional probability of error), since successive samples are almost independent; $r$ is larger for fast fading, since the length of error bursts is decreased.
A comparison between the throughput performance of GBN and SR is presented in Fig. 6. As expected, SR performs better than GBN, especially for fast fading. It is interesting to see how the Doppler frequency affects the performance. In GBN, clustering of errors has a beneficial effect, since isolated errors result in $m$ retransmissions anyway; therefore, GBN performs better in slow fading. On the other hand, the converse is true for SR, where the clustering of errors in the forward direction has no effect, whereas isolated errors in the feedback allow the system to exploit better the time-diversity feature.

A comparison between Fig. 6(a) and (b) shows that Rician fading yields better results than Rayleigh fading, as expected, since the propagation conditions are less severe. For Rician fading, the gain of SR over GBN is smaller. Also, SR is less sensitive than is GBN to the Rice factor.

Fig. 7 shows the throughput versus $F$ for various values of the round-trip delay, $m$, and the time-out, $t$. For slow fading, the performance is slightly improved by increasing $t$ (time diversity is better exploited); for fast fading, on the other hand, this leads to worse performance. Time diversity in GBN is beneficial as long as the increase in the value of $t$ allows the protocol to recover feedback errors, but does not cause too many packets to be retransmitted in a row; otherwise, the latter effect outweighs the former [5]. Also, for slow fading (long bursts), the sensitivity to $m$ and $t$ is much weaker than for fast fading (since the scale of time variations is expanded). In the presence of diversity, a similar behavior can be observed [Fig. 7(b)], with generally better performance.

Analogously, the dependence of the performance of SR on the time-diversity is investigated in Fig. 8. In this case, unlike for GBN, the throughput performance depends only on $\Delta = t - m$ [6]. Fig. 8 shows the performance for $\Delta = 0$ (i.e., no time diversity) and for $\Delta = \infty$, which corresponds to perfect feedback. Note that, in these two cases, the throughput is given by $\eta = (1 - \varepsilon)^2$ and $1 - \varepsilon$, respectively [6], and does not depend on the Doppler frequency. The performance for intermediate $\Delta$ will lie in between these curves, and will approach the perfect feedback performance for fast fading. In particular, small values of $\Delta$ will yield almost the same performance as $\Delta = \infty$ in fast fading, since bursts of errors are short and a small amount of time diversity is sufficient
to recover almost all errors; the converse is true for slow fading.

VI. CONCLUSIONS

In this paper, we studied the fading process, according to the commonly adopted nonrational model proposed by Jakar [22]. We found that, depending on the Doppler frequency and the data rate, the fading can be considered as “fast” or “slow.” The distinction between the two types of fading is not clear in the literature: often, in studies on packet transmission, slow fading denotes the situation in which all bits in a packet fade together, whereas fast fading denotes the situation in which each bit in the packet fades independently of the others.

This is too simplistic, and a more elaborate model for the relative dependence among segments of a message is desirable. What we proposed takes into account the correlation between different samples of the channel.

Although we used Jakar’s model (2) as a reference, most considerations do not depend on the exact shape of the fading spectrum, but rather on some general correlation properties. Our results depend primarily on the Doppler frequency $f_D$ and apply to other fading processes as well. Our evaluations, based on both theoretical considerations and practical examples, show that such a representation can be found, and can be approximated by means of channel models which are easy to study and have physical significance.

REFERENCES


**Michele Zorzi** (S’89–M’95), for a photograph and biography, see this issue, p. 436.

**Ramesh R. Rao** (SM’90) was born in Sindri, India, in 1958. He received the Bachelor’s degree with honors in electrical and electronics engineering from the University of Madras in 1980. He received the M.S. degree in 1982 and the Ph.D. degree in 1984, both from the University of Maryland, College Park. Since then, he has been on the faculty of the Department of Electrical and Computer Engineering at the University of California at San Diego (UCSD), La Jolla. His research interests include architectures, protocols, and performance analysis of computer and communication networks.

**Laurence B. Milstein** (S’66–M’68–SM’75–F’85) received the B.E.E. degree from the City College of New York in 1964 and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of Brooklyn, NY, in 1966 and 1968, respectively. From 1968 to 1974, he was employed by the Space and Communications Group of Hughes Aircraft Company, and from 1974 to 1976 he was a Member of the Department of Electrical and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Since 1976, he has been with the Department of Electrical and Computer Engineering, University of California at San Diego (UCSD), La Jolla, where he is a Professor and former Department Chairman working in the area of digital communication theory with special emphasis on spread-spectrum communication systems. He has also been a consultant to both government and industry in the areas of radar and communications.

Dr. Milstein was an Associate Editor for Communication Theory for the *IEEE TRANSACTIONS ON COMMUNICATIONS*, an Associate Editor for book reviews for the *IEEE TRANSACTIONS ON INFORMATION THEORY*, an Associate Technical Editor for the *IEEE Communications Magazine*, and is currently a Senior Editor for the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*. He was the Vice President For Technical Affairs of the IEEE Communications Society in 1990 and 1991, and has been a Member of the Board of Governors of both the IEEE Communications Society and the IEEE Information Theory Society. He is also a Member of Eta Kappa Nu and Tau Beta Pi.