Coding for the fading channel: a survey

Article *in* Signal Processing · July 2000

DOI: 10.1016/S0165-1684(00)00027-X · Source: DBLP

3 authors:

Ezio Biglieri
University Pompeu Fabra

Giuseppe Caire
Technische Universität Berlin

Giorgio Taricco
Politecnico di Torino

Signal Processing for Multimedia
J.S. Byrnes (Ed.)
IOS Press, 1999

Coding for the Fading Channel: a Survey

Ezio Biglieri and Giuseppe Caire and Giorgio Taricco

Dipartimento di Elettronica
Politecnico di Torino
Italy
`<name>@polito.it`

ABSTRACT. We consider the rationale behind the selection of coding schemes for wireless channels. Optimum coding schemes for this channel lead to the development of new criteria for code design, differing markedly from the Euclidean-distance criterion which is commonplace over the additive white Gaussian noise (AWGN) channel. For example, for a flat, slow-fading Rayleigh channel the code performance depends strongly, rather than on the minimum Euclidean distance of the code, on its minimum Hamming distance (the "code diversity"). If the channel model is not stationary, as happens for example in a mobile-radio communication system where it may fluctuate in time between the extremes of Rayleigh and AWGN, then a code designed to be optimum for a fixed channel model might perform poorly when the channel varies. Therefore, a code optimal for the AWGN channel may be actually suboptimum for a substantial fraction of time. In these conditions, antenna diversity with maximum-gain combining may prove useful: in fact, under fairly general conditions, a channel affected by fading can be turned into an AWGN channel by increasing the number of diversity branches. Another robust solution is based on bit interleaving, which yields a large diversity gain thanks to the choice of powerful convolutional codes coupled with a bit interleaver and the use of a suitable bit metric. An important feature of bit-interleaved coded modulation is that it lends itself quite naturally to "pragmatic" designs, i.e., to coding schemes that keep as their basic engine an off-the-shelf Viterbi decoder. Yet another solution is based on controlling the transmitted power so as to compensate for the attenuations due to fading.

1. Introduction: The fading channel

In the simplest communication channel model (the "additive white Gaussian channel", or AWGN) the received signal is assumed to be affected only by a constant attenuation and a constant delay. Digital transmission over radio channels often needs a more elaborate model, since it may be necessary to account for propagation vagaries, referred to as "fading," which affect the signal strength. These are connected with a propagation environment referred to as "multipath" and with the relative movement of transmitter and receiver, which causes time variations of the channel.

Multipath propagation occurs when the electromagnetic energy carrying the modulated signal propagates along more than one "path" connecting the transmitter to the receiver. Examples of such situations occur in indoor propagation, when the electromagnetic waves are perturbed by structures inside the building, and in terrestrial mobile radio, when multipath is caused by large fixed or moving objects (buildings, hills, cars, etc.). Two manifestations of channel time variations are delay spread and Doppler-frequency spread. A fading-channel classification can be based on these two parameters.

This work was supported by the Italian Space Agency (ASI)..

1.0.1. *Delay spread.* The signal components arriving from the various paths (direct and indirect) with different delays combine to produce a distorted version of the transmitted signal. To characterize by a single constant the various delays incurred by the signal traveling through the channel, we define a *delay spread* as the difference between the largest and the smallest among these delays. We say that this delay spread causes the two effects of *time dispersion* and *frequency-selective fading*.

Let B_x denote the bandwidth of the transmitted signal. If this is narrow enough so that the signal is not distorted, there is no frequency selectivity. As B_x increases, the distortion becomes increasingly noticeable. A measure of the signal bandwidth beyond which the distortion becomes relevant is usually given in terms of the so-called *coherence bandwidth* of the channel, denoted by B_c and defined as the inverse of the delay spread. The coherence bandwidth is the frequency separation at which two frequency components of the signal undergo independent attenuations. A signal with $B_x \gg B_c$ is subject to frequency-selective fading. More precisely, the envelope and phase of two unmodulated carriers at different frequencies will be markedly different if their frequency spacing exceeds B_c, so that the cross-correlation of the fading fluctuations of the two tones decreases toward zero. The term "frequency-selective fading" expresses this lack of correlation among different frequency components of the transmitted signal.

1.0.2. *Doppler-frequency spread.* When the receiver and the transmitter are in relative motion with constant radial speed, the received signal is subject to a constant frequency shift (the *Doppler shift*) proportional to this speed and to the carrier frequency. This Doppler effect, in conjunction with multipath propagation, causes *frequency dispersion* and *time-selective fading*. Frequency dispersion, in the form of an increase of the bandwidth occupancy of a signal, occurs when the channel changes its characteristics during signal propagation. Doppler-frequency spread is in a sense dual to delay spread.

The power spectrum of the signal received from each path is a sum of signals, each of which has a different frequency shift depending on its path. We have *frequency dispersion*. We define the "Doppler spread" as the difference between the largest and the smallest among the frequency shifts of the various paths.

A measure of the signal duration beyond which this distortion becomes relevant is given in terms of the so-called *coherence time* of the channel, denoted by T_c and defined as the inverse of the Doppler spread. Let T_x denote the duration of a transmitted pulse. If this is so short that during transmission the channel does not change its features appreciably, then the signal will be received undistorted. Its distortion becomes noticeable when T_x is well above T_c, the delay between two time components of the signal beyond which their attenuations become independent.

1.0.3. *Fading-channel classification.* From the previous discussion we have seen that the two quantities B_c and T_c describe how the channel behaves for the transmitted signal. Specifically,

- **(i)** If $B_x \ll B_c$, there is no frequency-selective fading, and hence no time dispersion. The channel transfer function looks constant, and the channel is called *flat* (or *non-selective*) in frequency.
- **(ii)** If $T_x \ll T_c$, there is no time-selective fading, and the channel is called *flat* (or *non-selective*) in time.

Qualitatively, the situation appears as shown in Fig. 1. The channel flat in t and f is not subject to fading, neither in time nor in frequency. The channel flat in time and selective in frequency is called the *intersymbol interference channel*. The channel flat in frequency is a good model for several terrestrial mobile radio channels, and most of the following will be devoted to its analysis. The selective channel, affected by fading both in time and in frequency, is not a good model for terrestrial mobile radio channels. It can be useful for

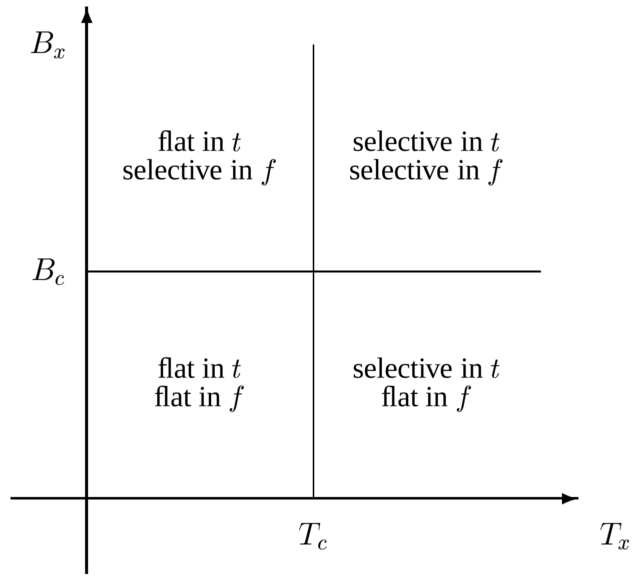

FIGURE 1. *Radio-channel classification.*

avionic communications, in which high speeds (and hence short coherence times) combine with long delays (and hence narrow coherence bandwidths) due to earth reflections.

2. Coding for the fading channel

Coding solutions for the fading channel should be selected by taking into account the distinctive features of the model used. Our goal here is to survey these solutions, by highlighting a number of issues that make code design for the fading channel differ from that for the AWGN channel. In this survey we examine, in particular, the effects of three features that make the fading channel differ from AWGN: namely, the fading channel is generally not memoryless (unless infinite-depth interleaving is assumed, an assumption that may not be realistic in several instances), has a signal-to-noise ratio which is a random variable rather than a constant, and finally the propagation vagaries may cause the channel model to vary with time, so that any model chosen may be able to represent the channel only for a fraction of the time.

2.1. *Speech vs. data: The delay issue*

A relevant factor in the choice of a coding scheme is the decoding delay that one may allow. For example, recently proposed, extremely powerful codes (the “Turbo Codes” of [4]) suffer from a considerable decoding delay, and hence their applicability is restricted.

Consider, for example, real-time speech transmission. Here, a strict decoding delay is imposed (e.g., 100 ms, at most [30]). In this case, the transmission of a code word may span only a few TDMA channel bursts, over which the channel fading is strongly correlated. Thus, a code word experiences only a few significant fading values, which makes the assumption of a memoryless channel, normally achieved by ideal or very long interleaving, no longer valid. On the contrary, with data traffic a large interleaving delay is tolerable, so that very effective coding techniques are available. For example, as we shall see, convolutional codes, bit interleaving, and high-level modulation (such as 8PSK or 16QAM) can be used. These techniques are generally referred to as Bit-Interleaved Coded Modulation (BICM) and have been extensively studied in [9, 10, 13] (Combination of BICM with the “turbo” idea has even been recently proposed, with very promising results. See [23–26]). Capacity calculations show that with large interleaving BICM performs as well as optimal coding over more complicated alphabets, and its complexity is much lower, so that the performance-complexity trade-off of BICM is very attractive. Moreover, capacity calculations [17] show that constant-power constant-rate transmission performs very close to optimal transmission schemes where power and rate are adapted dynamically to the channel conditions via a perfect feedback link. Then, with large interleaving and powerful coding, there is no need for implementing such complicated adaptive techniques and feedback links.

2.1.1. *Modeling the delay constraints.* The delay constraints can be easily taken into account when designing a coding scheme if a "block-fading" channel model is used. In this model, the fading process is about constant for a number of symbol intervals. On such a channel, a single code word may be transmitted after being split into several blocks, each suffering from a different attenuation, thus realizing an effective way of achieving diversity.

The "block-fading" channel model, introduced in [19, 30], is motivated by the fact that, in many mobile radio situations, the channel coherence time is much longer than one symbol interval, and hence several transmitted symbols are affected by the same fading value. Use of this channel model allows one to introduce a delay constraint for transmission, which is realistic whenever infinite-depth interleaving is not a reasonable assumption.

This model assumes that a code word of length $n = MN$ spans M blocks of length N (a group of M blocks will be referred to as a *frame*.) The value of the fading in each block is constant. M turns out to be a measure of the interleaving *delay* of the system. In fact, $M = 1$ corresponds to $N = n$, i.e., to no interleaving, while $M = n$ corresponds to $N = 1$, and hence to ideal interleaving. Thus, the results for different values of M illustrate the downside of nonideal interleaving. It should also be observed that the coding scheme implied by this channel model generalizes standard diversity techniques. In fact, the latter can be seen as a special case of coding for a block-fading channel on which repetition codes are used.

With no delay constraint, a code word can span an arbitrarily large number M of fading blocks. If this is the case then capacity, as derived in [17], is a good performance indicator. This applies, for example, to variable-rate systems (e.g., wireless data networks). On the other hand, most of today's mobile radio systems carry real-time speech (cellular telephony), for which constant-rate, constrained-delay transmission should be considered. In the latter case, that is, when each code word must be transmitted and decoded within a frame of $M < \infty$ blocks, *information outage rate*, rather than capacity, is the appropriate performance limit indicator. We shall not delve in this issue any further here, and the interested reader is referred to [15, 30].

2.2. *Diversity*

Receiver-diversity techniques have been known for a long time to improve the fading-channel quality. Recently, their synergy with coding has been extensively investigated in [37–39]. The standard approach to antenna diversity is based on the fact that, with several diversity branches, the probability that the signal will be simultaneously faded on all branches can be made small. The approach taken in [33, 37–39] is philosophically different, as it is based upon the observation that, under fairly general conditions, a channel affected by fading can be turned into an additive white Gaussian noise (AWGN) channel by increasing the number of diversity branches. Consequently, it can be expected (and it was indeed verified by analyses and simulations) that a coded modulation scheme designed to be optimal for the AWGN channel will perform asymptotically well also on a fading channel with diversity, at the only cost of an increased receiver complexity. An advantage of this solution is its robustness, since changes in the physical channel affect the reception very little.

This allows us to argue that the use of "Gaussian" codes along with diversity reception provides a solution to the problem of designing robust coding schemes for the mobile radio channel.

Recently, a considerable amount of work has been presented on *transmitter* diversity with coding [2, 16, 29, 34, 35]. Transmitter diversity may be a viable solution for the downlink of a TDMA system, where the receiver is constrained to be small and light (so that use of multiple antennas is out of the question).

2.3. *Multi-user detection: the challenge*

The design of coding schemes is further complicated when a multi-user environment is accounted for. The main problem here, and in general in communication systems that share channel resources, is the presence of multiple-access interference (MAI). This is generated by the fact that every user receives, besides the signal which is specifically directed to it, also some power from transmission to other users. This is true not only when CDMA is used, but also with space-division multiple access, in which intelligent antennas are directed towards the intended user. The earlier studies devoted to multi-user transmission simply neglected the presence of MAI. Typically, they were based on the naive assumption that, due to some version of the ubiquitous "Central Limit Theorem," signals adding up from a variety of users would coalesce to a process resembling Gaussian noise. Thus, the effect of MAI would be an increase of thermal noise, and any coding scheme designed to cope with the latter would still be optimal, or at least near-optimal, for multiuser systems.

Of late it was recognized that this assumption was groundless, and consequently several of the conclusions that it prompted were wrong. The central development of multi-user theory was the introduction of the optimum multi-user detector. Rather than demodulating each user separately and independently, it demodulates all of them simultaneously. A simple example should suffice to appreciate the extent of the improvement that can be achieved by optimum detection. In the presence of vanishingly small thermal noise, optimum detection would provide error-free transmission, while standard ("single-user") detection is affected by an error probability floor which increases with the number of users. Multi-user detection was born in the context of terrestrial cellular communication, and hence implicitly assumed an MAI-limited environment where thermal noise is negligible with respect to MAI (high-SNR condition). For this reason coding was seldom considered, and hence almost all multiuser detection schemes known from the literature are concerned with symbol-by-symbol decisions.

2.4. *Unequal error protection*

In some analog source coding applications, like speech or video compression, the sensitivity of the source decoder to errors in the coded symbols is typically not uniform. The quality of the reconstructed analog signal is rather insensitive to errors affecting certain classes of bits, while it degrades sharply when errors affect other classes. This happens, for example, when analog source coding is based on some form of hierarchical coding, where a relatively small number of bits carries the "fundamental information" and a larger number of bits carries the "details," as occurs in the case of the MPEG2 standard.

Assuming that the source encoder produces frames of binary coded symbols, each frame can be partitioned into classes of symbols of different "importance" (i.e., of different sensitivity). Then, it is apparent that the best coding strategy aims at achieving lower BER levels for the important classes while admitting higher BER levels for the unimportant ones. This feature is referred to as unequal error protection (UEP). On the contrary, codes for which the BER is (almost) independent of the position of the information symbols are referred to as equal error protection (EEP) codes.

An efficient method for achieving UEP with Turbo Codes was recently studied in [11]. The key point is to match a non-uniform puncturing pattern to the interleaver of the Turbo-encoder in order to create locally low-rate Turbo Codes for the important symbols, and locally high-rate Turbo Codes for the unimportant symbols. In this way, we can achieve several protection levels while keeping the total code rate constant. On the decoding side, we only need to "depuncture" the received sequence by inserting zeros at the punctured positions. Then a single Turbo-decoder can handle different code rates, both equal-error-protection Turbo Codes and UEP Turbo Codes.

2.5. *The frequency-flat, slow Rayleigh-fading channel*

This channel model assumes that the duration of a modulated symbol is much greater than the delay spread caused by the multipath propagation. If this occurs, then all frequency components in the transmitted signal are affected by the same random attenuation and phase shift, and the channel is frequency-flat. If, in addition, the channel varies very slowly with respect to the symbol duration, then the fading $R(t)\exp[j\Theta(t)]$ remains approximately constant during the transmission of one symbol. If this does not occur the fading process is called *fast*.

The assumption of non-selectivity allows us to model fading as a process affecting the transmitted signal in a multiplicative form. The assumption of slow fading allows us to model this process as a constant random variable during each symbol interval. In conclusion, if $x(t)$ denotes the complex envelope of the modulated signal transmitted during the interval $(0, T)$, then the complex envelope of the signal received at the output of a channel affected by slow, flat fading and additive white Gaussian noise can be expressed in the form

$$r(t) = Re^{j\Theta}x(t) + n(t), \tag{1}$$

where $n(t)$ is a complex Gaussian noise and $Re^{j\Theta}$ is a Gaussian random variable, with R having a Rice or Rayleigh pdf and unit second moment, i.e., $E[R^2] = 1$.

If we can further assume that the fading is so slow that we can estimate the phase shift Θ with sufficient accuracy, and hence compensate for it, then coherent detection is feasible. Thus, model (1) can be further simplified to

$$r(t) = Rx(t) + n(t). \tag{2}$$

It should be immediately apparent that with this simple model of the fading channel the only difference with respect to an AWGN channel resides in the fact that R, instead of being a constant attenuation, is now a random variable whose value affects the amplitude, and hence the power, of the received signal. Assume finally that the value taken by R is known at the receiver. We describe this situation by saying that we have *perfect* CSI. Channel state information can be obtained, for example, by inserting a pilot tone in a notch of the spectrum of the transmitted signal, and by assuming that the signal is faded exactly in the same way as this tone.

Detection with perfect CSI can be performed exactly in the same way as for the AWGN channel. In fact, the constellation shape is perfectly known, as is the attenuation incurred by the signal. The optimum decision rule in this case consists of minimizing the Euclidean distance

$$\int_0^T [r(t) - Rx(t)]^2\, dt \quad \text{or} \quad |\mathbf{r} - R\mathbf{x}|^2 \tag{3}$$

with respect to the possible transmitted real signals $x(t)$ (or vectors $\mathbf{x}$).

A consequence of this fact is that the error probability with perfect CSI and coherent demodulation of signals affected by frequency-flat, slow fading can be evaluated as follows. We first compute the error probability $P(e \mid R)$ obtained by assuming R constant in model (2), then we take the expectation of $P(e \mid R)$ with respect to the random variable R. The calculation of $P(e \mid R)$ is performed as if the channel were AWGN, but with the energy $\mathcal{E}$ changed into $R^2\mathcal{E}$. Notice finally that the assumptions of noiseless channel-state information and a noiseless phase-shift estimate make the values of $P(e)$ thus obtained as representing a limiting performance.

Consider now the error probabilities that we would obtain with binary signals without coding (see [5] for a more general treatment). For two signals with common energy $\mathcal{E}$ and

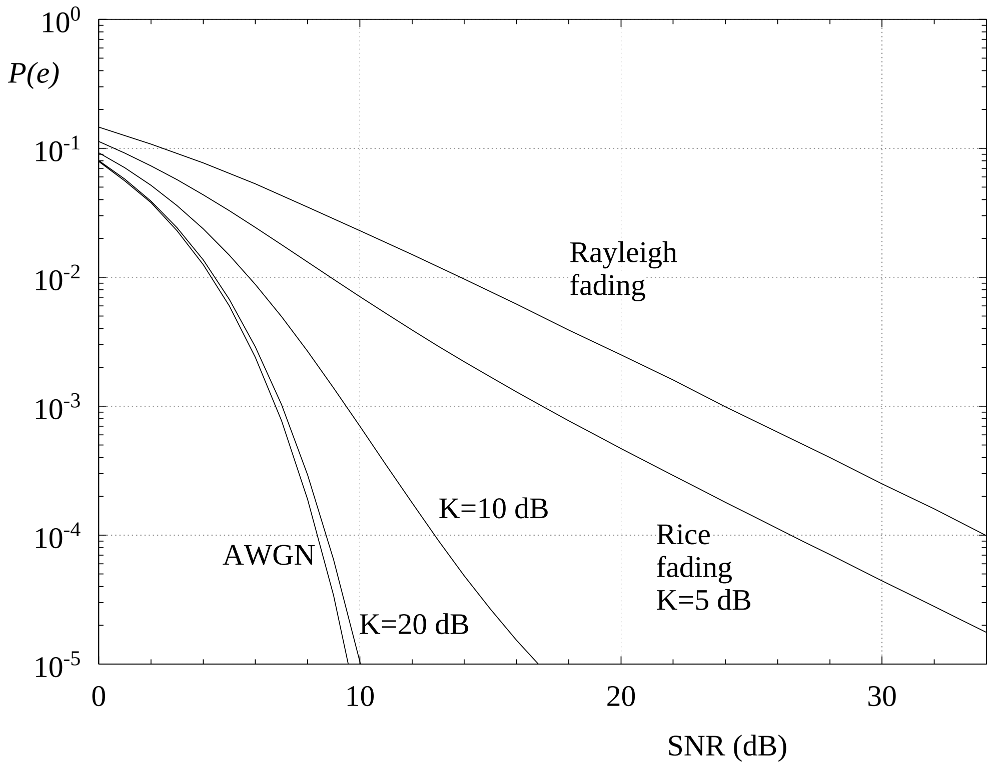

FIGURE 2. *Error probabilities of binary antipodal transmission over the Gaussian channel and over Rayleigh and Rice fading channels.*

correlation coefficient $\rho = (\mathbf{x}, \widehat{\mathbf{x}})/\mathcal{E}$ we have, for Rayleigh fading and perfect channel-state information,

$$P(e) = \frac{1}{2}\left(1 - \sqrt{\frac{(1-\rho)\mathcal{E}/2N_0}{1 + (1-\rho)\mathcal{E}/2N_0}}\right). \tag{4}$$

In the absence of CSI, one could take a decision rule consisting of minimizing

$$\int_0^T [r(t) - x(t)]^2 \, dt \quad \text{or} \quad |\mathbf{r} - \mathbf{x}|^2. \tag{5}$$

However, with constant envelope signals ($|\mathbf{x}|$ constant), the error probabilities obtained with (3) and (5) coincide, because

$$\begin{aligned} P(\mathbf{x} \to \widehat{\mathbf{x}}) &= P(|\mathbf{r} - R\widehat{\mathbf{x}}|^2 < |\mathbf{r} - R\mathbf{x}|^2) \\ &= P(2R(\mathbf{r}, \mathbf{x} - \widehat{\mathbf{x}}) < 0) \\ &= P((\mathbf{r}, \mathbf{x} - \widehat{\mathbf{x}}) < 0) \end{aligned}$$

and hence CSI is completely represented by the phase Θ. Fig. 2 compares error probabilities of binary antipodal transmission over the Gaussian channel with those over the Rayleigh and Rice fading channel, where K denotes the "Rice factor" of the latter [3, Chap. 13]. It is seen that the loss with increasing error probability is considerable. As we shall see in a moment, coding can compensate for a substantial amount of this loss.

2.6. *Our survey*

In the following we shall survey a few important issues in coding for the fading channel. The model we assume here is that of a channel affected by flat, slow fading and additive noise. Optimum coding schemes for this channel model lead to the development of new criteria for code design (Section 3). If the channel model is not stationary, as happens, for example, in a mobile-radio communication system, then a code designed to be optimum for a fixed channel model might perform poorly when the channel varies. Therefore, a code optimal for the AWGN channel may be actually suboptimum for a substantial fraction of time. In these conditions, antenna diversity with maximum-gain combining may prove useful. In fact, under fairly general conditions, a channel affected by fading can be turned into an AWGN channel by increasing the number of diversity branches (Section 4.1). Another robust solution is based on bit interleaving, which yields a large diversity gain thanks to the choice of powerful convolutional codes coupled with a bit interleaver and the use of a suitable bit metric (Section 4.2). Yet another solution is based on controlling the transmitted power so as to compensate for the attenuations due to fading (Section 4.3).

3. Code-design criteria

A standard code-design criterion, when soft decoding is chosen, is to choose coding schemes that maximize their minimum Euclidean distance. This is, of course, correct on the Gaussian channel with high SNR (although not when the SNR is very low: see [32]), and is often accepted, *faute de mieux*, on channels that deviate little from the Gaussian model (e.g., channels with a moderate amount of intersymbol interference). However, the Euclidean-distance criterion should be outright rejected over the Rayleigh fading channel. In fact, analysis of coding for the Rayleigh fading channel proves that Hamming distance (also called “code diversity” in this context) plays the central role here.

Assume transmission of a coded sequence $\mathcal{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ where the components of $\mathcal{X}$ are signal vectors selected from a constellation $\mathcal{S}$. We do not distinguish here among block or convolutional codes (with soft decoding), or block- or trellis-coded modulation.

3.1. *No delay constraint: Infinite-depth interleaving*

We also assume for the moment infinite-depth interleaving, which makes the fading random variables affecting the various symbols $\mathbf{x}_k$ independent. Hence we write, for the components of the received sequence $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$,

$$\mathbf{r}_k = R_k \mathbf{x}_k + \mathbf{n}_k, \tag{6}$$

where the R_k are independent and, under the assumption that the noise is white, the RV’s n_k are also independent.

Coherent detection of the coded sequence, with the assumption of perfect channel-state information, is based upon the search for the coded sequence $\mathcal{X}$ that minimizes the distance

$$\sum_{k=1}^{N} |\mathbf{r}_k - R_k \mathbf{x}_k|^2. \tag{7}$$

The pairwise error probability can be upper bounded in this case as [31], [3, Chap.13]

$$P\{\mathcal{X} \to \widehat{\mathcal{X}}\} \leq \prod_{k \in \mathcal{K}} \frac{1}{1 + |\mathbf{x}_k - \widehat{\mathbf{x}}_k|^2 / 4N_0} \tag{8}$$

where $\mathcal{K}$ is the set of indices k such that $\mathbf{x}_k \neq \widehat{\mathbf{x}}_k$.

3.1.1. *An example.*

For illustration purposes, let us compute the Chernoff upper bound to the word error probability of a block code with rate R_c. Assume that binary antipodal modulation is used, with waveforms of energies $\mathcal{E}$, and that the demodulation is coherent with perfect CSI. Observe that for $\widehat{\mathbf{x}}_k \neq \mathbf{x}_k$ we have

$$|\mathbf{x}_k - \widehat{\mathbf{x}}_k|^2 = 4\mathcal{E} = 4R_c\mathcal{E}_b,$$

where $\mathcal{E}_b$ denotes the average energy per bit. For two code words $\mathcal{X}$, $\widehat{\mathcal{X}}$ at Hamming distance $d_H(\mathcal{X}, \widehat{\mathcal{X}})$ we have

$$P\{\mathcal{X} \to \widehat{\mathcal{X}}\} \leq \left(\frac{1}{1 + R_c \mathcal{E}_b / N_0}\right)^{d_H(\mathcal{X}, \widehat{\mathcal{X}})}$$

and hence, for a linear code,

$$P(e) = P(e \mid \mathcal{X}) \leq \sum_{w \in \mathcal{W}} \left(\frac{1}{1 + R_c \mathcal{E}_b / N_0}\right)^{w},$$

where $\mathcal{W}$ denotes the set of nonzero Hamming weights of the code, considered with their multiplicities. It can be seen that for high enough signal-to-noise ratio the dominant term in the expression of $P(e)$ is the one with exponent $d_{\min}$, the minimum Hamming distance of the code. □

By recalling the above calculation, the fact that the probability of error decreases inversely with the signal-to-noise ratio raised to power $d_{\min}$ can be expressed by saying that we have introduced a *code diversity* $d_{\min}$.

We may further upper bound the pairwise error probability by defining the set $\mathcal{K}$ of indices k for which $\mathbf{x}_k \neq \widehat{\mathbf{x}}_k$, and writing

$$P\{\mathcal{X} \to \widehat{\mathcal{X}}\} \leq \prod_{k \in \mathcal{K}} \frac{1}{|\mathbf{x}_k - \widehat{\mathbf{x}}_k|^2 / 4N_0} = \frac{1}{[\delta^2(\mathcal{X}, \widehat{\mathcal{X}}) / 4N_0]^{d_H(\mathcal{X}, \widehat{\mathcal{X}})}} \tag{9}$$

(which is close to the true Chernoff bound for small enough N_0). Here

$$\delta^2(\mathcal{X}, \widehat{\mathcal{X}}) = \left[\prod_{k \in \mathcal{K}} |\mathbf{x}_k - \widehat{\mathbf{x}}_k|^2\right]^{1/d_H(\mathcal{X}, \widehat{\mathcal{X}})}$$

is the geometric mean of the non-zero squared Euclidean distances between the components of $\mathcal{X}$, $\widehat{\mathcal{X}}$. The latter result shows the important fact that the error probability is (approximately) inversely proportional to the *product* of the squared Euclidean distances between the components of $\mathbf{x}$, $\widehat{\mathbf{x}}$ that differ and, to a more relevant extent, to a power of the signal-to-noise ratio whose exponent is the Hamming distance between $\mathcal{X}$ and $\widehat{\mathcal{X}}$.

Further, we know from the results referring to block codes, convolutional codes, and coded modulation that the union bound to error probability for a coded system can be obtained by summing up the pairwise error probabilities associated with all the different "error events." For small noise spectral density N_0, i.e., for high signal-to-noise ratios, a few equal terms will dominate the union bound. These correspond to error events with the smallest value of the Hamming distance $d_H(\mathcal{X}, \widehat{\mathcal{X}})$. We denote this quantity by L_c to stress the fact, to be discussed soon, that it reflects a diversity residing in the code. We have

$$P\{\mathcal{X} \to \widehat{\mathcal{X}}\} \lessapprox \frac{\nu}{[\delta^2(\mathcal{X}, \widehat{\mathcal{X}}) / 4N_0]^{L_c}} \tag{10}$$

where ν is the number of dominant error events. For error events with the same Hamming distance, the values taken by $\delta^2(\mathcal{X}, \widehat{\mathcal{X}})$ and by ν are also of importance. This observation may be used to design coding schemes for the Rayleigh fading channel. Here, no role is played by the Euclidean distance, which is the central parameter used in the design of coding schemes for the AWGN channel.

For uncoded systems ($n = 1$), the results above hold with the positions $L_c = 1$ and $\delta^2(\mathcal{X}, \widehat{\mathcal{X}}) = |\mathbf{x} - \widehat{\mathbf{x}}|^2$, which shows that the error probability decreases as N_0. A similar result could be obtained for maximal-ratio combining in a system with diversity L_c. This explains the name of this parameter. In this context, the various diversity schemes may be seen as implementations of the simplest among the coding schemes, the repetition code, which provides a diversity equal to the number of diversity branches [31].

From the discussion above, we have learned that over the perfectly-interleaved Rayleigh fading channel the choice of a coding scheme should be based on the maximization of the code diversity, i.e., the minimum Hamming distance among pairs of error events. Since for the Gaussian channel code diversity does not play the same central role, coding schemes optimized for the Gaussian channel are likely to be suboptimum for the Rayleigh channel.

3.2. *Introducing delay constraints: The block-fading channel*

The above analysis holds, *mutatis mutandis*, for the block-fading channel: it suffices in this case to interpret the variables $\mathbf{x}_k$ as *blocks of symbols*, rather than symbols. In this situation, it should not come as a surprise (and can in fact be shown rigorously, see [20, 22]) that the relevant criterion becomes the *block-Hamming* distance, i.e., the number of *blocks* in which two code words differ. An application of the Singleton Bound shows that the maximum block-Hamming distance achievable on an M-block fading channel is limited by

$$D \leq 1 + \left\lfloor M \left(1 - \frac{R}{\log_2 |\mathcal{S}|} \right) \right\rfloor$$

where $|\mathcal{S}|$ is the size of the signal set $\mathcal{S}$ and R is the code rate, expressed in bits/symbol. Note that binary signal sets ($|\mathcal{S}| = 2$) are not effective in this case, so that codes constructed over high-level alphabets should be considered [20, 22].

For a deeper analysis of the relationship between code diversity and code rate, see [27,28].

4. Robust coding schemes

The design procedure described in the section above, and consisting of adapting the coding scheme to the channel, may suffer from a basic weakness. If the channel model is not stationary, as it is, for example, in a mobile-radio environment where it fluctuates in time between the extremes of Rayleigh and AWGN, then a code designed to be optimum for a fixed channel model might perform poorly when the channel varies. Therefore, a code optimal for the AWGN channel may be actually suboptimum for a substantial fraction of time. An alternative solution consists of doing the opposite, i.e., *matching the channel to the coding scheme*: the latter is still designed for a Gaussian channel, while the former is transformed from a Rayleigh-fading channel (say) into a Gaussian one. Here we shall examine three such robust solutions, the first based on antenna diversity, the second on bit-interleaving, and the third on power control.

4.1. *Antenna diversity*

Fig. 3 shows the block diagram of the transmission scheme with fading. A source of co-channel interference is also added for completeness. Our initial assumptions, valid in the following unless otherwise stated, are [37–39]:

- PSK modulation
- M independent diversity branches whose signal-to-noise ratio is inversely proportional to M (this assumption is made in order to disregard the SNR increase that actually occurs when multiple receive elements are used).
- Flat, independent Rayleigh fading channel.
- Coherent detection with perfect channel-state information.
- Synchronous diversity branches.
- Independent co-channel interference, and a single interferer.

The codes examined are the following:

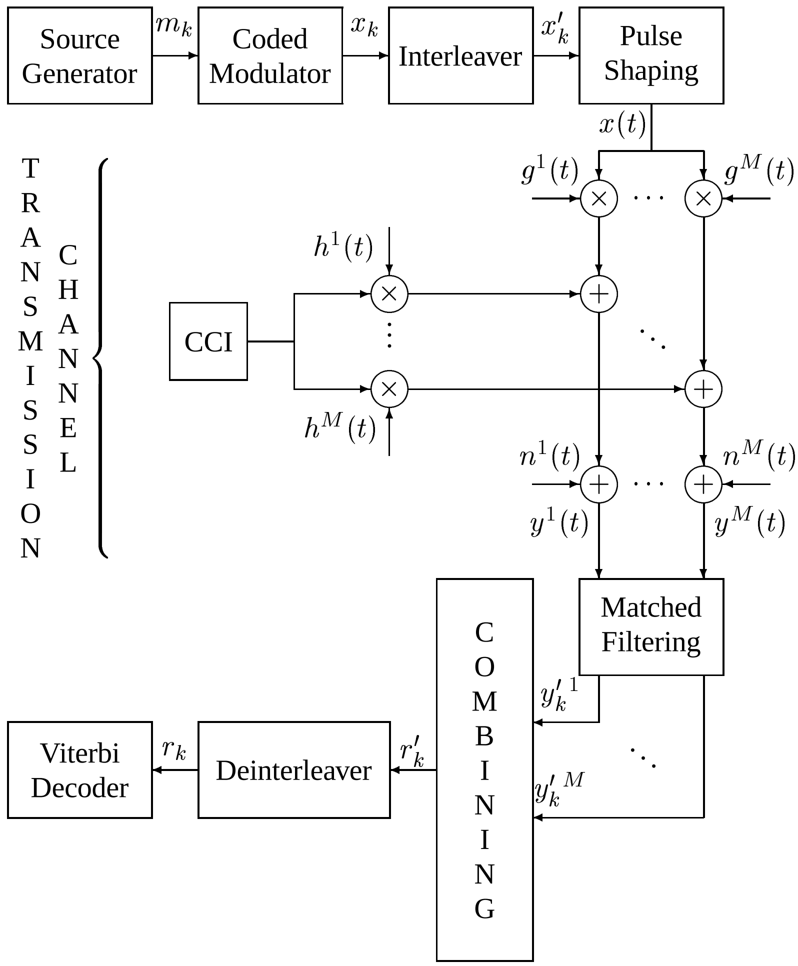

FIGURE 3. *Block diagram of the transmission scheme.*

- *J4*: 4-state, rate-2/3 TCM scheme based on 8-PSK and optimized for Rayleigh-fading channels [18].
- *U4:* 4-state rate-2/3 TCM scheme based on 8-PSK and optimized for the Gaussian channel.
- *U8:* Same as above, with 8 states.
- *Q64:* "Pragmatic" concatenation of the "best" rate-1/2 64-state convolutional code with 4-PSK modulator and Gray mapping [40].

Fig. 4 compares the performance of U4 and J4 (two TCM schemes with the same complexity) over a Rayleigh-fading channel with M-branch diversity. It is seen that, as M increases, the performance of U4 comes closer and closer to that of J4. Similar results hold for correlated fading. Even for moderate correlation J4 loses its edge on U4, and for M as low as 4 U4 performs better than J4 [37]. The effect of diversity is more marked when the code used is weaker. As an example, two-antenna diversity provides a gain of 10 dB at BER=10^{-6} when U8 is used, and of 2.5 dB when Q64 is used [37]. The assumption of branch independence, although important, is not critical. In effect, [37] shows that branch correlation coefficients as large as .5 degrade system BER only slightly. The complexity introduced by diversity can be traded for delay. As shown in [37], in some cases diversity makes interleaving less necessary, so that a lower interleaving depth (and consequently a lower overall delay) can be compensated by an increase of M.

When differential or pilot-tone, rather than coherent, detection is used [38], a BER-floor occurs which can be reduced by introducing diversity. As for the effect of co-channel interference, even its BER-floor is reduced as M increases (although for its elimination multi-user detectors should be employed). This shows that antenna diversity with maximal-ratio combining is highly instrumental in making the fading channel closer to Gaussian.

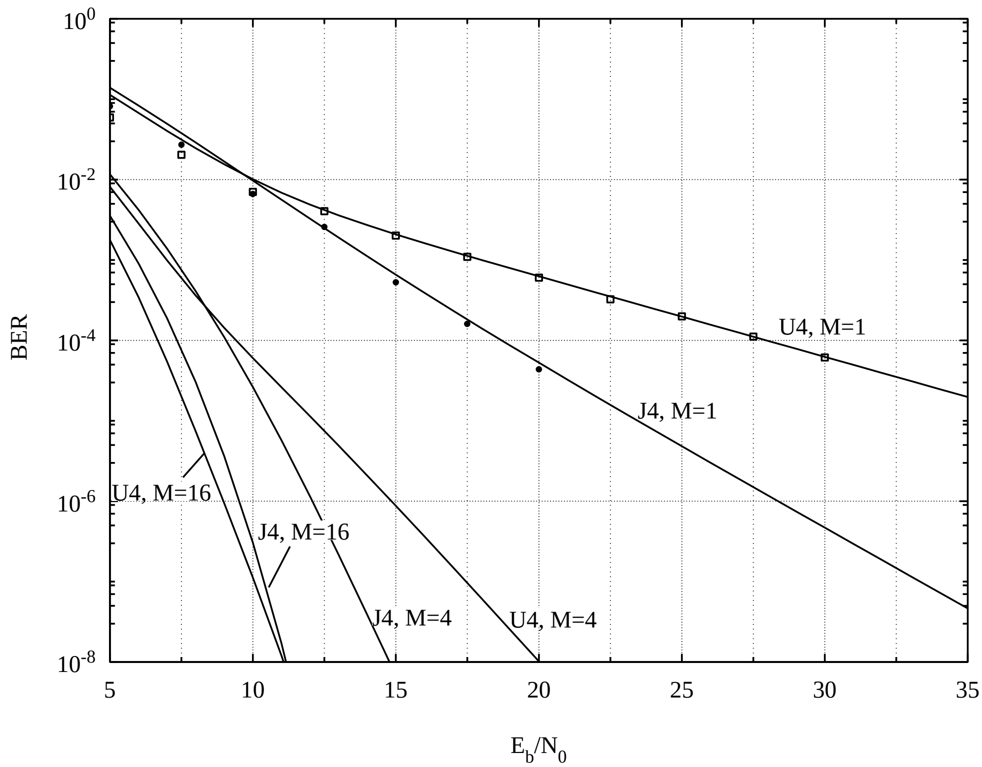

FIGURE 4. *Effect of antenna diversity on the performance of 4-state TCM schemes over the flat, independent Rayleigh-fading channel. J4 is optimum for the Rayleigh channel, while U4 is optimum for the Gaussian channel.*

4.2. *Bit-interleaved coded modulation*

Ever since 1982, when Ungerboeck published his landmark paper on trellis-coded modulation [36], it has been generally accepted that modulation and coding should be combined in a single entity for improved performance. Several results followed this line of thought, as documented by a considerable body of work aptly summarized and referenced in [18] (see also [6, Chap. 10]). Under the assumption that the symbols were interleaved with a depth exceeding the coherence time of the fading process, new codes were designed for the fading channel so as to maximize their diversity. This implied in particular that parallel transitions should be avoided in the code, and that any increase in diversity would be obtained by increasing the constraint length of the code. One should also observe that for non-Ungerboeck systems, i.e., those separating modulation and coding with binary modulation, Hamming distance is proportional to Euclidean distance, and hence a system optimized for the additive white Gaussian channel is also optimum for the Rayleigh fading channel.

A notable departure from Ungerboeck's paradigm was the core of [40]. Schemes were designed in which coded modulation is generated by pairing an M-ary signal set with a binary convolutional code with the largest minimum free Hamming distance. Decoding was achieved by designing a metric aimed at keeping as their basic engine an off-the-shelf Viterbi decoder for the *de facto* standard, 64-state rate-$1/2$ convolutional code. This implied giving up the joint decoder/demodulator in favor of two separate entities.

Based on the latter concept, Zehavi [42] first recognized that the code diversity, and hence the reliability of coded modulation over a Rayleigh fading channel, could be further improved. Zehavi's idea was to make the code diversity equal to the smallest number of distinct *bits* (rather than *channel symbols*) along any error event. This is achieved by bit-wise interleaving at the encoder output, and by using an appropriate soft-decision bit metric as an input to the Viterbi decoder. For different approaches to the problem of designing coded modulation schemes for the fading channels see [8].

One of Zehavi's findings, rather surprising *a priori*, was that on some channels there is a downside to combining demodulation and decoding. This prompted the investigation whose results are presented in a comprehensive fashion in [13] (see also [1]).

An advantage of this solution is its robustness, since changes in the physical channel affect the reception very little. Thus, it provides good performance with a fading channel as well as with an AWGN channel (and, consequently, with a Rice fading channel, which can be seen as intermediate between these two).

Recently, a very promising combination of BICM with the concept of turbo-decoding [23–26] has been discussed and analyzed.

4.3. *Power control*

Observation of (2) shows that what makes this Rayleigh fading channel differ from AWGN is the fact that R is a random variable rather than a constant attenuation. Consequently, if this variability of R could be compensated for, an AWGN would be obtained. This compensation can be achieved in principle if channel-state information is available to the transmitter, which consequently can modulate its power according to the channel fluctuations.

Consider the simplest such strategy. The flat independent fading channel with coherent detection yields the received signal (2). Assume that the channel state information R is known at the transmitter front-end, that is, the transmitter knows the value of R during the transmission (this assumption obviously requires that R is changing very slowly). Under these conditions, assume that the transmitted signal in an interval with length T is

$$x(t) = \sigma s(t), \tag{11}$$

where $s(t)$ has unit energy (equal-energy basic waveform) and σ is chosen under a given optimality criterion.

One possible such criterion (constant error probability over each symbol) requires that

$$\sigma = R^{-1}. \tag{12}$$

This way, the channel is transformed into an equivalent additive white Gaussian noise channel. The error probability is the same as if we had transmitted s over a channel whose only effect is the addition of n to the transmitted signal. The average transmitted power per symbol is then

$$\mathrm{E}[x^2(t)] = \mathrm{E}[1/\rho^2], \tag{13}$$

which might diverge.

This technique ("channel inversion") is simple to implement, since the encoder and decoder are designed for the AWGN channel, independent of the fading statistics. For instance, it is common in spread-spectrum systems with near-far interference imbalances. However, it may suffer from a large capacity penalty. For example, in Rayleigh fading the capacity is zero.

To avoid divergence of the average power (or an inordinately large value thereof) a possible strategy is the following. Choose

$$\sigma = \begin{cases} R^{-1} & \text{if } R > R_0 \\ R_0^{-1} & \text{otherwise.} \end{cases} \tag{14}$$

By choosing the value of the threshold R_0 appropriately we trade off a decrease of the average power value for an increase of error probability. The average power value is now

$$(1-p)\,\frac{1}{R_0^2} + p\,\mathrm{E}[1/R^2 \mid R > R_0], \tag{15}$$

where $p = \mathrm{P}[R > R_0]$. For an information-theoretic analysis of power-control techniques for the fading channel, see [12].

5. Conclusions

This review was aimed at illustrating some concepts that make the design of codes for the fading channel differ markedly from the same task applied to the Gaussian channel. In particular, we have examined the design of "fading codes," i.e., coding schemes which maximize the Hamming, rather than the Euclidean, distance, the interaction of antenna diversity with coding (which makes the channel more Gaussian), the effect of separating coding from modulation in favor of a more robust coding scheme, and the effect of transmitter-power control. The issue of optimality as contrasted to robustness was also discussed to some extent.

References

[1] S. A. Al-Semari and T. Fuja, "Bit interleaved I-Q TCM," *ISITA '96*, Victoria, B.C., September 17–20, 1996.

[2] H. Asakura and T. Matsumoto, "Cooperative signal reception and down-link beamforming in cellular mobile communications," *Proceeding of ISIT'97*, Ulm, Germany, 1997.

[3] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. New York: Plenum Press, 1998.

[4] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, Vol. 44, No. 10, pp. 1261–1271, October 1996.

[5] E. Biglieri, G. Caire, and G. Taricco, "Error probability over fading channels: A unified approach," *European Transactions on Telecommunications*, January 1998.

[6] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*, New York: Macmillan, 1991.

[7] J. Boutros and E. Viterbo, "Rotated multidimensional constellations for Rayleigh fading channels," *1996 IEEE Information Theory Workshop*, Haifa, Israel, June 9-13, 1996.

[8] J. Boutros, E. Viterbo, C. Rastello, and J.-C. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 502–518, March 1996.

[9] G. Caire, E. Biglieri and G. Taricco, "Bit-interleaved coded modulation," Proceedings of *ICC '97*, Montreal, Canada, June 1997.

[10] G. Caire, G. Taricco and E. Biglieri, "Capacity of bit-interleaved channels," *IEE Electronics Letters*, Vol. 32, No. 12, pp. 1060 – 1061, June 1996.

[11] G. Caire and G. Lechner, "Turbo-codes with unequal error protection," *IEE Electronics Letters*, Vol. 32 No. 7, pp. 629, March 1996.

[12] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, Submitted for publication, December 1997.

[13] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, May 1998.

[14] G. Caire, G. Taricco and E. Biglieri, "Optimal power control for minimum outage rate in wireless communications," submitted to *ICC '98*, Aug. 1997.

[15] D. Falconer and G. Foschini, "Theory of minimum mean-square-error QAM systems employing decision feedback equalization," *Bell System Tech. J.* 52, pp. 1821 – 1849, 1973.

[16] M. Fitz, J. Grimm and J. Krogmeier, "Results on code design for transmitter diversity in fading," *Proceedings of ISIT'97*, Ulm, Germany, 1997.

[17] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, to appear, 1997.

[18] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. New York: Kluwer Academic Publishers, 1994.

[19] G. Kaplan and S. Shamai (Shitz), "Error probabilities for the block-fading Gaussian channel," *A.E.U.*, Vol. 49, No. 4, pp. 192 – 205, 1995.

[20] R. Knopp, *Coding and Multiple Access over Fading Channels*, Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 1997.

[21] R. Knopp and G. Caire, "Simple power-controlled modulation schemes for fixed rate transmission over fading channels," to be submitted to *IEEE Trans. Commun.*, 1997.

[22] R. Knopp, P. A. Humblet, "Maximizing diversity on block fading channels," *Proceedings of ICC '97*, Montréal, Canada, 8 – 12 June, 1997.

[23] X. Li and J. A. Ritchey, "Bit-interleaved coded modulation with iterative decoding—Approaching turbo-TCM performance without code concatenation," *Proc. CISS'98*, Princeton, NJ, March 18–20, 1998.

[24] X. Li and J. A. Ritchey, "Bit-interleaved coded modulation with iterative decoding using soft feedback," *Submitted for publication*, 1998.

[25] X. Li and J. A. Ritchey, "Low-complexity MAP decoding for trellis coded modulation with bit interleaving," *Proc. Globecom'97 Communication Theory Mini-Conference*, Phoenix, AZ, Nov. 1997, pp. 28–32.

[26] X. Li and J. A. Ritchey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Communications Letters*, Vol. 1, No. 6, pp. 169–171, November 1997.

[27] E. Malkamäki and H. Leib, "Coded diversity schemes on block fading Rayleigh channels," *IEEE Int. Conf. on Universal Personal Communications, ICUPC'97*, San Diego, CA, October 1997.

[28] E. Malkamäki and H. Leib, "Coded diversity on block fading channels," *Submitted to IEEE Trans. Inform. Theory*, 1997.

[29] A. Narula, M. D. Trott and G. Wornell, “Information theoretic analysis of multiple antenna transmitter diversity,” submitted to *IEEE Trans. Inform. Theory,* Nov. 1996.

[30] L. Ozarow, S. Shamai, and A. D. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Trans. Vehic. Tech.*, Vol. 43, No. 2, May 1994.

[31] N. Seshadri and C.-E. W. Sundberg, “Coded modulations for fading channels—An overview,” *European Trans. Telecomm.*, Vol. ET-4, No. 3, pp. 309–324, May-June 1993.

[32] M. Steiner, “The strong simplex conjecture is false,” *IEEE Trans. Inform. Theory*, pp. 721–731, May 1994.

[33] G. Taricco, E. Biglieri, and G. Caire, “Impact of channel-state information on coded transmission over fading channels with diversity reception,” *submitted for publication*, 1998.

[34] V. Tarokh, N. Seshadri and R. Calderbank “Space-time codes for high data rate wireless communication: Performance criteria,” Proceedings of *ICC’97*, Montréal, Canada, June 1997.

[35] V. Tarokh, A. Naguib, N. Seshadri and R. Calderbank “Space-Time codes for high data rate wireless communication: Mismatch analysis,” Proceedings of *ICC’97*, Montréal, Canada, June 1997.

[36] G. Ungerboeck, “Channel coding with multilevel/phase signals”, *IEEE Trans. Inform. Th.*, vol. IT-28, pp. 56–67, Jan. 1982.

[37] J. Ventura-Traveset, G. Caire, E. Biglieri, and G. Taricco, “Impact of diversity reception on fading channels with coded modulation—Part I: Coherent detection,” *IEEE Trans. Commun.*, Vol. 45, No. 5, May 1997.

[38] J. Ventura-Traveset, G. Caire, E. Biglieri, and G. Taricco, “Impact of diversity reception on fading channels with coded modulation—Part II: Differential block detection,” *IEEE Trans. Commun.*, Vol. 45, No. 6, June 1997.

[39] J. Ventura-Traveset, G. Caire, E. Biglieri, and G. Taricco, “Impact of diversity reception on fading channels with coded modulation—Part III: Co-channel interference,” *IEEE Trans. Commun.*, Vol. 45, No. 7, July 1997.

[40] A. J. Viterbi, J. K. Wolf, E. Zehavi, and R. Padovani, “A pragmatic approach to trellis-coded modulation,” *IEEE Communications Magazine*, vol. 27, n. 7, pp. 11–19, 1989.

[41] J. Yang and S. Roy, “Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback,” *IEEE Trans. Inform. Theory,* Vol. 40, No. 5, pp. 1334 – 1347, Sept. 1994.

[42] E. Zehavi, “8-PSK Trellis Codes for a Rayleigh channel,” *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873-884, May 1992.