ELSEVIER

Signal Processing 80 (2000) 1135–1148

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

Coding for the fading channel: a survey

Ezio Biglieri, Giuseppe Caire, Giorgio Taricco*,[1]

Dipartimento di Elettronica, Politecnico di Torino, c/so duca degli Abruzzi 24, 10129 Turnio, Italy

Received 28 November 1997

Abstract

We consider the design of coding schemes for a channel affected by flat, slow fading and additive noise. Specifically, by using the "block-fading" channel model, we allow delay constraints to be taken into account. Optimum coding schemes for this channel model 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. In fact, 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 it 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. © 2000 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Wir betrachten den Entwurf von Kodes für einen Kanal, der durch flaches, langsames Fading und additives Rauschen gestört ist. Speziell durch die Benutzung des "Block-Fading" Kanalmodells orlauben wir die Berücksichtigung von Verzögerungsbedingungen. Optimale Kodierschemata für dieses Kanalmodell führen zur Entwicklung neuer Kriterien für den Kodeentwurf, die sich deutlich vom Kriterium der Euklidischen Distanz unterscheiden, welches allgemein für Kanäle mit additivem, weißem Gaußschem Rauschen (AWGN) Verwendung findet. Tatsächlich hängt die Leistungsfähigkeit des Kodes eher von seiner minimalen Hamming Distanz (der "code diversity") ab, als von der minimalen Euklidschen Distanz des Kodes. Ist das Kanalmodell nicht stationär, wie es beispielsweise im Mobilfunk der Fall ist, wo es zwischen den Extremfällen Rayleigh und AWGN schwanken kann, dann kann ein für ein bestimmtes Kanalmodell optimaler Kode ein schlechtes Verhalten zeigen, wenn sich der Kanal ändert. Deshalb kann ein für AWGN Kanäle optimaler Kode tatsächlich während einer beträchtlichen Zeitspanne suboptimal sein. Unter diesen Bedingungen kann sich Antennen-Diversität mit "maximum-gain combining" als nützlich erweisen: tatsächlich kann unter weitestgehend

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

* Corresponding author. Tel.: + 39-011-56-40-00; fax: + 39-011-564-40-99.

E-mail address: taricco@polito.it (G. Taricco).

PII: S0165-1684(00)00027-X

allgemeinen Bedingungen ein durch Fading betroffener Kanal durch Erhöhung der Anzahl der Diversitätspfade in einen AWGN-Kanal umgewandelt werden. Eine weitere robste Lösung basiert auf Interleaving, die wiederum einen großen Diversitätsgewinn dank der Wahl leistungsstarker Faltungskodes, verbunden mit Bit-Interleaving und der Verwendung einer geeigneten Bit-Metrik liefert. Ein wichtiges Merkmal von bit-interleaved kodierter Modulation ist, daß sie sich fast natürlich für pragmatischen Entwurf eignet, d.h. für Kodes, die als ihr wesentliches Werkzeug einen gewöhnlichen Viterbi-Dekoder behalten. Noch eine andere Lösung basiert auf der Kontrolle der Sendeleistung, um so die Dämpfungen durch Fading auszugleichen.

Résumé

Dans ce papier nous considérons la conception de systèmes de codage pour des canaux avec évanouissements lents, non sélectifs en fréquence et bruit additif. Particulièrement, en utilisant un modèle de canal de type "évanouissement par blocs" les contraintes de retard peuvent être prises en compte. Le système de codage optimal pour ce type de canal implique le développement de nouveaux critères de conception qui différent clairement du critère de distance Euclidienne utilisé dans le classique canal fixe avec bruit additif Gaussien. En fait, la performance du code dépends fortement de la distance minimum de Hamming (Diversité de code) plutôt que de la distance Euclidienne minimum. Si le canal est non stationnaire comme dans le cas mobile (allant du canal de Rayleigh au canal avec bruit additif Gaussien), alors, un code adapté au canal fixe peut avo ir de mauvaises performances sur canal mobile. Un code optimal adapté au canal fixe sera ainsi sous-optimal dans le cas mobile. Dans ces conditions, la diversité d'antenne peut être utile : en fait, sous certaines conditions peu contraignantes, un canal avec évanouissements peut tendre vers un canal fixe avec bruit additif Gaussien si le nombre d'antennes est grand. Une autre solution efficace est l'entrelacement des données qui amène un gain substantiel du au couple codeur convolutionnel/entrelaceur. Une importante propriété des couples entrelaceur/codeur convolutionnel est l'utilisation pragmatique de décodeurs de Viterbi standards. Enfin, Une autre solution est le contrôle de la puissance émise afin de compenser les évanouissements.

Keywords: Fading channel; Coded modulation; Pragmatic code; Capacity of fading channel; Bit interleaving; Multiuser detection; Unequal error protection

1. Introduction: general considerations

The increasing practical relevance of digital mobile-radio transmission systems has led of late to a great deal of work devoted to coding for fading channels. For these channels the paradigms developed for the Gaussian channel may not be valid anymore, and a fresh look at the design philosophies is called for. Specifically, system designer choices were often driven by their knowledge of coding over the AWGN channel: consequently, they tried to apply to radio channels solutions that were far from optimum on channels where nonlinearities, Doppler shifts, fading, shadowing, and interference from other users make the channel far from Gaussian.

A considerable body of work has been reversing this "Gaussian" perspective, and it is now being widely recognized that coding solutions for the fading channel should be selected by taking into account the distinctive features of the fading channel. The goal of this paper is to survey this work, by listing 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 make the channel model to vary with time, so that any chosen model may be able to represent the channel only for a fraction of the time.

1.1. 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, the recently proposed, extremely powerful codes (the "Turbo Codes" [3]) 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 [28]). 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 [8,10,12]. 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 [18] 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.

1.1.1. 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, and thus realizing an effective way of achieving diversity.

The "block-fading" channel model, introduced in [21,28], 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 [18], 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 [16,28].

1.2. 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 [34–36]. 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 [34–36] 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 indeed 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 [2,17,27,31,32]. 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 multiple antennas are not implementable).

1.3. 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 multi-user 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 a 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.

1.4. 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" like 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 [9]. 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 constant the total code rate. On the decoding side, all what we need is to "depuncture" the received sequence by inserting zeros at the punctured positions. Then, a single turbo-decoder can handle different code rates, equal-error-protection turbo codes and UEP turbo codes.

1.5. The frequency flat, slow Rayleigh-fading channel

Before discussing all this, we present in a tutorial fashion some results on the capacity of the Rayleigh fading channels: these results show the importance of coding on this channel, and the relevance of obtaining channel state information (CSI) in the demodulation process. 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 the symbol duration, then the fading $R(t)\exp[\mathrm{j}\Theta(t)]$ 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 the fading as a process affecting the transmitted signal in a multiplicative form. The assumption of a 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

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

where $n(t)$ is a complex Gaussian noise, and $R\mathrm{e}^{\mathrm{j}\Theta}$ 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.[2] 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 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

$$\int_0^T [r(t) - Rx(t)]^2\,\mathrm{d}t = |\boldsymbol{r} - R\boldsymbol{x}|^2 \tag{3}$$

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

[2] This represents part of the CSI described above. A receiver recovering the phase shift Θ has achieved part of its goal of obtaining the CSI, and in certain cases, like with binary PSK, it has achieved it all.

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 $P(e\,|\,R)$ obtained by assuming R constant in model (2), then we take the expectation of $P(e\,|\,R)$, with respect to the random variable R. The calculation of $P(e\,|\,R)$ is performed as if the channel were AWGN, but with the energy $\mathscr{E}$ changed into $R^2\mathscr{E}$. Notice finally that the assumptions of a 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 [4] for a more general treatment). For two signals with common energy $\mathscr{E}$ and correlation coefficient $\rho = (\boldsymbol{x},\hat{\boldsymbol{x}})/\mathscr{E}$ we have, for Rayleigh fading and perfect channel-state information

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

This can be upper bounded, for large signal-to-noise ratios, by

$$P(e) \leqslant \frac{2}{1-\rho}\,\frac{1}{\mathscr{E}/N_0}. \tag{5}$$

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

$$\int_0^T [r(t) - x(t)]^2\,\mathrm{d}t = |\boldsymbol{r} - \boldsymbol{x}|^2. \tag{6}$$

However, with constant envelope signals ($|\boldsymbol{x}|$ constant), the error probability obtained with (3) and (6) coincide because

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

and hence CSI is completely represented by the phase Θ. Fig. 1 compares error probabilities over the Gaussian channel with those over the Rayleigh-fading channel with perfect CSI. It is seen

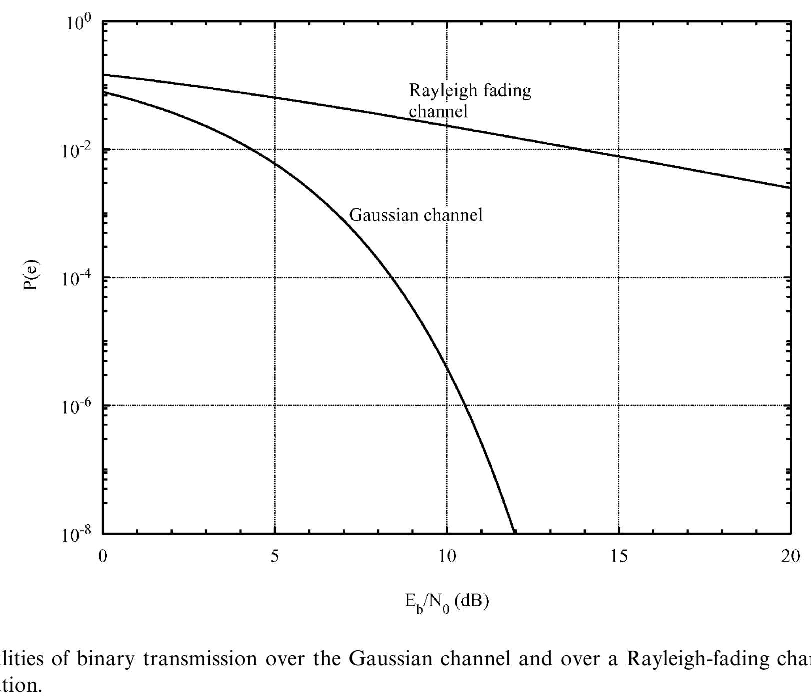

Fig. 1. Error probabilities of binary transmission over the Gaussian channel and over a Rayleigh-fading channel with and without channel-state information.

that the loss in error probability is considerable. As we shall see in a moment, coding can compensate for a substantial amount of this loss.

1.6. Our survey

In this paper we survey a few important issues in coding for the fading channel. The model we assume 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 it 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).

2. Impact of coding: a capacity analysis

The effects of coding will be analyzed here by computing the fading channel capacity with binary inputs. These results will show that the capacity loss with respect to the AWGN channel is considerably lower than that of uncoded transmission. This predicts the effectiveness of coding.

The channel model is described in Fig. 2. Here, x is the channel input, a the channel state, y the channel output, and $\hat{a}$ the estimate of the channel state. We assume that y depends only on a, x, and not on $\hat{a}$.

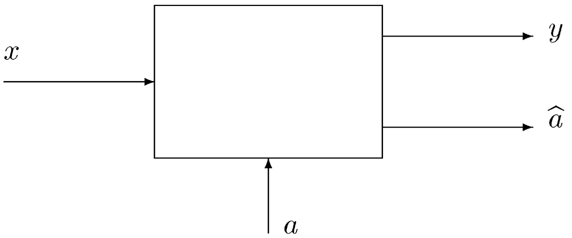

Fig. 2. Fading channel model: x is the channel input, a the channel state, y the channel output, and $\hat{a}$ the estimate of the channel state.

The information we obtain on x from the observation of y and $\hat{a}$ is given by

$$I(x;y,\hat{a}) = \int_x \int_y \int_{\hat{a}} p(x,y,\hat{a}) \log_2 \frac{p(x\,|\,y,\hat{a})}{p(x)} \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\hat{a},$$

where $p(\cdot)$ are probabilities or probability density functions. The channel capacity is given by

$$C = \max I(x;y,\hat{a}),$$

where the maximum is to be taken over the distribution of x. The calculation of this maximum can be avoided by choosing for $p(x)$ the uniform distribution $1/M$ (M the number of channel symbols). Under suitable symmetry assumptions this choice provides the actual capacity. By making the simplifying assumption that transmission is binary, the symmetry condition is

$$p(y\,|\,a,-x) = p(-y\,|\,a,x)$$

which yields, with $p(x) = \frac{1}{2}$,

$$C = 1 + E_{\hat{a}}\left[\int_y p(y\,|\,\hat{a},1) \log_2 \frac{p(y|\hat{a},1)}{p(y|\hat{a},1) + p(y|\hat{a},-1)} \,\mathrm{d}y\right]. \quad (7)$$

By observing that

$$p(\hat{a}) = \int_a p(\hat{a}\,|\,a) p(a) \,\mathrm{d}a,$$

we have

$$p(y\,|\,\hat{a},x) = \int_a p(a\,|\,\hat{a}) p(y\,|\,a,x) \,\mathrm{d}a. \quad (8)$$

We have two limiting special cases here:

Perfect channel-state information: This occurs when $\hat{a} = a$, so that (7) yields

$$C = E_a[C(a)], \tag{9}$$

where $C(a)$ is the "conditional capacity" of the channel when the latter is in state a:

$$C(a) = 1 + \int_y p(y|a,1)\log_2 \frac{p(y|a,1)}{p(y|a,1) + p(y|a,-1)}\,\mathrm{d}y. \tag{10}$$

No channel-state information: In this case

$$p(\hat{a}|a) = p(\hat{a}),$$

so that from (8) we have

$$C = 1 + \int_y p(y|1)\log_2 \frac{p(y|1)}{p(y|1) + p(y|-1)}\,\mathrm{d}y. \tag{11}$$

If we compare (Fig. 3) the channel capacities of a Rayleigh-fading channel with and without CSI with that of a Gaussian channel, we see that the loss in capacity due to the fading is much smaller than the loss in terms of error probability. This shows that coding for the fading channel is actually highly beneficial.

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 [30]), 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 $\mathscr{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ where the components of $\mathscr{X}$ are signal vectors selected from a constellation

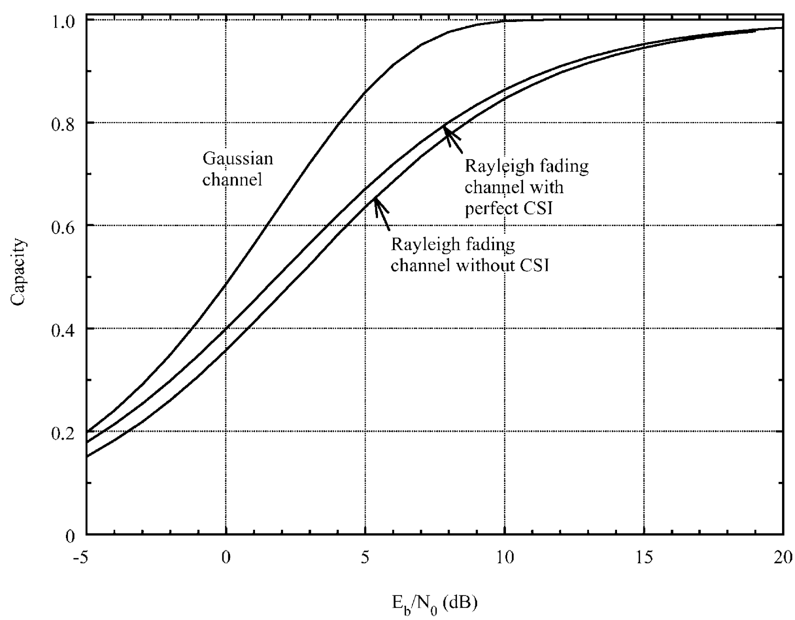

Fig. 3. Capacity of binary transmission over the Gaussian channel and over a Rayleigh-fading channel with and without channel-state information.

$\mathscr{S}$. 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 $\boldsymbol{x}_k$ to be independent. Hence we write, for the components of the received sequence $(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_n)$:

$$\boldsymbol{r}_k = R_k \boldsymbol{x}_k + \boldsymbol{n}_k, \tag{12}$$

where the R_k are independent, and, under the assumption that the noise is white, the RVs 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 $\mathscr{X}$ that minimizes the distance

$$\sum_{k=1}^{N} |\boldsymbol{r}_k - R_k \boldsymbol{x}_k|^2. \tag{13}$$

The pairwise error probability can be upper bounded in this case as [29]

$$P\{\mathscr{X} \to \hat{\mathscr{X}}\} \leqslant \prod_{k \in \mathscr{K}} \frac{1}{1 + |\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k|^2 / 4N_0}, \tag{14}$$

where $\mathscr{K}$ is the set of indices k such that $\boldsymbol{x}_k \neq \hat{\boldsymbol{x}}_k$.

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 $\mathscr{E}$, and that the demodulation is coherent with perfect CSI. Observe that for $\hat{\boldsymbol{x}}_k \neq \boldsymbol{x}_k$ we have

$$|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k|^2 = 4\mathscr{E} = 4R_c\mathscr{E}_b,$$

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

$$P\{\mathscr{X} \to \hat{\mathscr{X}}\} \leqslant \left(\frac{1}{1 + R_c\mathscr{E}_b/N_0}\right)^{H(\mathscr{X}, \hat{\mathscr{X}})}$$

and hence, for a linear code,

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

where $\mathscr{W}$ 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 writing

$$P\{\mathscr{X} \to \hat{\mathscr{X}}\} \leqslant \prod_{k \in \mathscr{K}} \frac{1}{|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k|^2 / 4N_0} = \frac{1}{[\delta^2(\mathscr{X}, \hat{\mathscr{X}})/4N_0]^{H(\mathscr{X}, \hat{\mathscr{X}})}} \tag{15}$$

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

$$\delta^2(\mathscr{X}, \hat{\mathscr{X}}) = \left[\prod_{k \in \mathscr{K}} |\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k|^2\right]^{H(\mathscr{X}, \hat{\mathscr{X}})}$$

is the geometric mean of the non-zero squared Euclidean distances between the components of $\mathscr{X}$, $\hat{\mathscr{X}}$. 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 $\boldsymbol{x}$, $\hat{\boldsymbol{x}}$ that differ, and, to a more relevant extent, to a power of the signal-to-noise ratio whose exponent is the Hamming distance between $\mathscr{X}$ and $\hat{\mathscr{X}}$.

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 $H(\mathscr{X}, \hat{\mathscr{X}})$. 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

$$P\{\mathscr{X} \to \hat{\mathscr{X}}\} \lessapprox \frac{v}{[\delta^2(\mathscr{X}, \hat{\mathscr{X}})/4N_0]^{L_c}}, \tag{16}$$

where v is the number of dominant error events. For error events with the same Hamming distance, the values taken by $\delta^2(\mathscr{X}, \hat{\mathscr{X}})$ and by v 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 $\delta^2(\mathscr{X}, \hat{\mathscr{X}}) = |\boldsymbol{x} - \hat{\boldsymbol{x}}|^2$, 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 [29].

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 $\boldsymbol{x}_k$ 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 [22,24]) that the relevant criterion becomes the *block-Hamming* distance, i.e., the number of *blocks* in which two code words differ. An application of Singleton Bound shows that the maximum block-Hamming distance achievable in this case is limited by

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

where $|\mathscr{S}|$ is the size of the signal set $\mathscr{S}$ and R is the code rate, expressed in bit/symbol. Note that binary signal sets ($|\mathscr{S}| = 2$) are not effective in this case, so that codes constructed over high-level alphabets should be considered [22,24].

For a deeper analysis of the relationship between code diversity and code rate, see [25,26].

4. Robust coding schemes

The design procedure described in the section above, and consisting of adapting the C/M 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. 4 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 [34–36]:

- PSK modulation.
- M independent diversity branches whose signal-to-noise ratio is inversely proportional to M (this

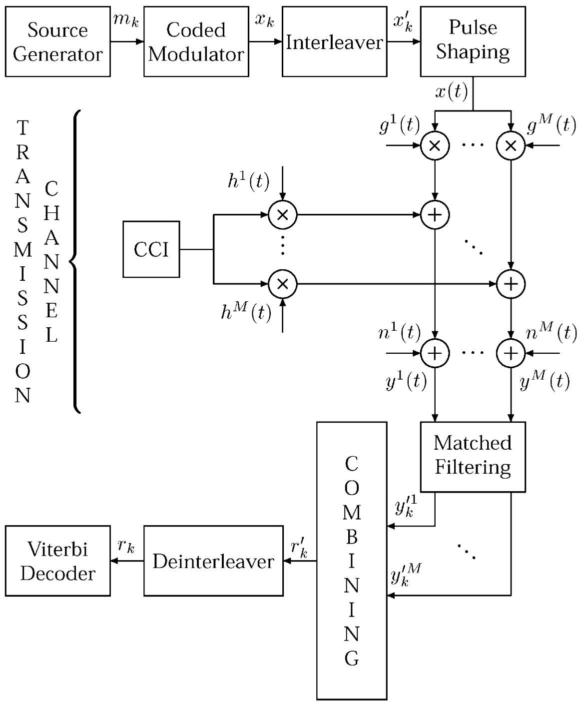

Fig. 4. Block diagram of the transmission scheme.

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:

J4. Four-state, rate-$\frac{2}{3}$ TCM scheme based on 8-PSK and optimized for Rayleigh-fading channels [20].

U4. Four-state rate-$\frac{2}{3}$ 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-$\frac{1}{2}$ 64-state convolutional code with 4-PSK modulator and Gray mapping [37].

Fig. 5 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

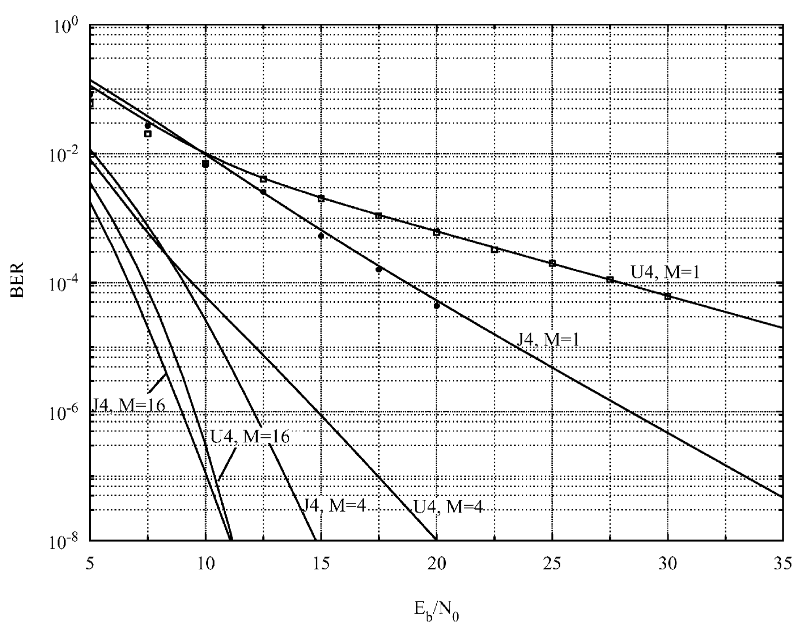

Fig. 5. Effect of antenna diversity on the performance of four-state TCM schemes over the flat, independent Rayleigh-fading channel. J4 is optimum for the Rayleigh channel, while U4 is optimum for the Gaussian channel.

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 [34]. 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 [34]. The assumption of branch independence, although important, is not critical: in effect, [34] shows that branch correlations as large as 5 degrade system BER only slightly. The complexity introduced by diversity can be traded for delay: as shown in [34], 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 [35], 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 increased (although for its elimination multiuser detectors should be employed). This shows that antenna diversity with maximal-ratio combining is highly instrumental in making the fading channel closer to Gaussian.

4.2. Bit-interleaved coded modulation

Ever since 1982, when Ungerboeck published his landmark paper on trellis-coded modulation [33], 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 [20] (see also [5, Chapter 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 [37]. 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-$\frac{1}{2}$ convolutional code. This implied giving up the joint decoder/demodulator in favor of two separate entities.

Based on the latter concept, Zehavi [39] 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 [7].

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 [12] (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 the latter two).

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

$$x(t) = \sigma s(t), \tag{17}$$

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

$$\sigma = R^{-1}. \tag{18}$$

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

$$E[x^2(t)] = E[1/\rho^2], \tag{19}$$

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

$$\sigma = \begin{cases} R^{-1} & \text{if } R > R_0, \\ R_0^{-1} & \text{otherwise.} \end{cases} \tag{20}$$

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

$$(1-p)\frac{1}{R_0^2} + pE[1/R^2 | R > R_0], \tag{21}$$

where $p = \mathrm{P}[R > R_0]$. For an information-theoretical analysis of power-control techniques for the fading channel, see [11].

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., C/M 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 C/M scheme, and the effect of transmitter-power control. The issue of optimality as contrasted to robustness was also discussed to some extent.

6. Further reading

The following references are also of interest to the reader: [6,13,15,19,23,38].

References

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

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

[3] C. Berrou, A. Glavieux, Near optimum error correcting coding and decoding: Turbo-codes, IEEE Trans. Commun. 44 (10) (October 1996) 1261–1271.

[4] E. Biglieri, G. Caire, G. Taricco, Error probabilitiy over fading channels: a unified approach, European Trans. Telecommun. (January 1998).

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

[6] J. Boutros, E. Viterbo, Rotated multidimensional constellations for Rayleigh fading channels, 1996 IEEE Information Theory Workshop, Haifa, Israel, 9–13 June 1996.
[7] J. Boutros, E. Viterbo, C. Rastello, J.-C. Belfiore, Good lattice constellations for both Rayleigh fading and Gaussian channels, IEEE Trans. Inform. Theory 42 (2) (March 1996) 502–518.
[8] G. Caire, E. Biglieri, G. Taricco, Bit-interleaved coded modulation, Proceedings of ICC '97, Montreal, Canada, June 1997.
[9] G. Caire, G. Lechner, Turbo-codes with unequal error protection, IEE Electron. Lett. 32 (7) (March 1996) 629.
[10] G. Caire, G. Taricco, E. Biglieri, Capacity of bit-interleaved channels, IEE Electron. Lett. 32 (12) (June 1996) 1060–1061.
[11] G. Caire, G. Taricco, E. Biglieri, Optimum power control over fading channels, IEEE Trans. Inform. Theory (December 1997), submitted for publication.
[12] G. Caire, G. Taricco, E. Biglieri, Bit-interleaved coded modulation, IEEE Trans. Inform. Theory (1997) to be published.
[13] G. Caire, G. Taricco, E. Biglieri, Optimal power control for minimum outage rate in wireless communications, ICC '98, August 1997, submitted for publication.
[15] T. Ericson, A Gaussian channel with slow fading, IEEE Trans. Inform. Theory 16 (May 1970) 353–355.
[16] D. Falconer, G. Foschini, Theory of minimum mean-square-error QAM systems employing decision feedback equalization, Bell System Technol. J. 52 (1973) 1821–1849.
[17] M. Fitz, J. Grimm, J. Krogmeier, Results on code design for transmitter diversity in fading, Proceedings of ISIT'97, Ulm, Germany, 1997.
[18] A. Goldsmith, P. Varaiya, Capacity of fading channels with channel side information, IEEE Trans. Inform. Theory (1997), to appear.
[19] J. Hagenauer, Zur Kanalkapazität bei Nachrichtenkanälen mit Fading und gebündelten Fehlern, Arch. Fuer Electron. und Uebertragungstech. 34 (6) (June 1980) 229–237.
[20] S.H. Jamali, T. Le-Ngoc, Coded-Modulation Techniques for Fading Channels, Kluwer Academic Publishers, New York, 1994.
[21] G. Kaplan, S. Shamai (Shitz), Error probabilities for the block-fading Gaussian channel, Arch. fuer Electron. und Uebertragungstech. 49 (4) (1995) 192–205.
[22] R. Knopp, Coding and multiple access over fading channels, Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 1997.
[23] R. Knopp, G. Caire, Simple power-controlled modulation schemes for fixed rate transmission over fading channels, IEEE Trans. Commun. (1997) to be submitted.
[24] R. Knopp, P.A. Humblet, Maximizing diversity on block fading channels, Proceedings of ICC '97, Montréal, Canada, 8–12 June, 1997.
[25] E. Malkamäki, H. Leib, Coded diversity schemes on block fading Rayleigh channels, IEEE International Conference on Universal Personal Communications, ICUPC'97, San Diego, CA, October 1997.
[26] E. Malkamäki, H. Leib, Coded diversity on block fading channels, IEEE Trans. Inform. Theory (1997), submitted for publication.
[27] A. Narula, M.D. Trott, G. Wornell, Information theoretic analysis of multiple antenna transmitter diversity, IEEE Trans. Inform. Theory (November 1996), submitted for publication.
[28] L. Ozarow, S. Shamai, A.D. Wyner, Information theoretic considerations for cellular mobile radio, IEEE Trans. Vehicle Technol. 43 (2) (May 1994).
[29] N. Seshadri, C.-E.W. Sundberg, Coded modulations for fading channels – an overview, European Trans. Telecomm. ET-4 (3) (May-June 1993) 309–324.
[30] M. Steiner, The strong simplex conjecture is false, IEEE Trans. Inform. Theory (May 1994) 721–731.
[31] V. Tarokh, A. Naguib, N. Seshadri, R. Calderbank, Space-Time codes for high data rate wireless communication: mismatch analysis, Proceedings of ICC'97, Montréal, Canada, June 1997.
[32] V. Tarokh, N. Seshadri, R. Calderbank, Space-time codes for high data rate wireless communication: performance criteria, Proceedings of ICC'97, Montréal, Canada, June 1997.
[33] G. Ungerboeck, Channel coding with multilevel-phase signals, IEEE Trans. Inform. Theory IT-28 (January 1982) 56–67.
[34] J. Ventura-Traveset, G. Caire, E. Biglieri, G. Taricco, Impact of diversity reception on fading channels with coded modulation – Part I: coherent detection, IEEE Trans. Commun. 45 (5) (May 1997).
[35] J. Ventura-Traveset, G. Caire, E. Biglieri, G. Taricco, Impact of diversity reception on fading channels with coded modulation – Part II: differential block detection, IEEE Trans. Commun. 45 (6) (June 1997).
[36] J. Ventura-Traveset, G. Caire, E. Biglieri, G. Taricco, Impact of diversity reception on fading channels with coded modulation – Part III: co-channel interference, IEEE Trans. Commun. 45 (7) (July 1997).
[37] A.J. Viterbi, J.K. Wolf, E. Zehavi, R. Padovani, A pragmatic approach to trellis-coded modulation, IEEE Commun. Mag. 27 (7) (1989) 11–19.
[38] J. Yang, S. Roy, Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback, IEEE Trans. Inform. Theory 40 (5) (September 1994) 1334–1347.
[39] E. Zehavi, 8-PSK Trellis Codes for a Rayleigh channel, IEEE Trans. Commun. 40 (5) (May 1992) 873–884.