Adaptive Trellis-Coded Multiple-Phase-Shift-Keying for Rayleigh Fading Channels

Siavash M. Alamouti and Samir Kallel
UBC Department of Electrical Engineering

CICSR-TR92-018
Adaptive Trellis-Coded Multiple-Phase-Shift-Keying for Rayleigh Fading Channels* 

Siavash M. Alamouti    Samir Kallel

Department of Electrical Engineering
University of British Columbia
Vancouver, B.C.
V6T 1Z4 Canada
Telephone: (604) 822 4985
Facsimile: (604) 822 5949
Email: siavash@ee.ubc.ca or kallel@ee.ubc.ca

* This research was supported in part by the "National Sciences and Engineering Council of Canada, and a B.C. Science Council GREAT award."
Abstract—An adaptive scheme for trellis-coded modulation of MPSK signals called adaptive trellis-coded multiple-phase-shift-keying (ATCMPSK) is proposed for slowly Rayleigh fading channels. The adaptive scheme employs a slightly modified rate 1/2 convolutional encoder and the corresponding Viterbi decoder to realize a family of codes of different rates which are employed according to channel conditions. During poor channel conditions, trellis-coded QPSK (TCQPSK) together with repetition schemes are employed. As channel conditions improve, higher rate schemes such as trellis-coded 16PSK are used. An interleaving/deinterleaving method suitable for the adaptive scheme is proposed. Theoretical bounds for the error performance and an exact expression for the throughput of the proposed adaptive scheme are derived, and are compared against simulation results. Simulations have been performed to measure the performance of the scheme for different parameters and some nonideal conditions. It is shown that ATCMPSK results in considerable improvement in bit-error-rate (BER) performance of MPSK signals. Under ideal conditions, gains in the range of 3 – 20 dB are achieved over conventional fixed rate pragmatic trellis-coded schemes.
I. INTRODUCTION

Freedom of mobility is the latest trend in the evolution of digital communication systems. This trend has portrayed itself in the implementation of digital cellular systems, and recent proposals for a universal digital portable communication system [1] – [4]. The objective of such a system is to remove the barriers which limit the mobility of the users at both ends of a communication link.

Radio spectrum is a scarce resource. Therefore, one of the most important objectives in the design of a digital portable radio system or a digital cellular system is the efficient usage of the available spectrum to accommodate the ever increasing demand. This must be done with no sacrifice in power or transmission quality. Power consumption must be minimized in order to limit the size of the portable units and to ensure safety from electromagnetic radiation hazard. To minimize the required power without any sacrifice in bit-error-rate (BER) performance or the required bandwidth, error control techniques can be applied. Trellis-coded modulation [5], [6] which is based on combining the functions of coding and modulation has been widely recognized as a powerful error control technique suitable for applications in mobile communications [7], [8].

The radio link for either a portable or a vehicular unit can be characterized by time-varying multipath fading. Traditional methods suggest the use of fixed rate codes for the transmission of speech or data on a time varying mobile channel. Fixed rate codes, however, fail to explore the time varying nature of the mobile radio channel. Therefore, in order to keep the performance at a desirable level, they are designed for average or worse channel conditions. In order to explore the time-varying nature of the radio link, adaptive schemes can be employed [9] – [11]. In this paper we propose an adaptive scheme called adaptive trellis-coded multiple-phase-shift-keying (ATCMPSK). The ATCMPSK scheme is pragmatic in terms of its design and implementation. A rate 1/2 convolutional encoder, and a slightly modified Viterbi decoder are used to construct
a family of pragmatic trellis codes [12] which are then employed according to channel conditions in an adaptive manner. During good channel conditions, more information is sent using high rate pragmatic trellis codes. As channel conditions become worse, lower rate trellis codes and repetition codes are applied. We have devised a suitable mapping scheme, and interleaving method that suit the adaptive nature of the scheme. We have shown that considerable gains in BER can be achieved using our proposed scheme. Furthermore, there are provisions made in the universal portable digital communication system proposed in [1] for the hand set and base units to: measure the radio link quality; coordinate and agree on the channel to be used for each call; and to be able to identify each other’s radio transmission. The high level of communication and coordination between the transmitting and receiving units facilitates the use of the adaptive scheme into the proposed system.

In our theoretical analysis, we have assumed symbol by symbol adaptation according to CSI at the start of each transmission. However, some simulations have been carried out that take into account the effect of block adaptation. In our model, both the transmitter and the receiver can “sense” any change in channel conditions at discrete instants of symbol transmission. We do not deal with the question of characteristics of the feedback channel. The rate of transmission in the feedback channel effectively reduces the overall throughput of the system. However, as it will be shown, the degradation in throughput performance due to the transmission in the feedback channel required for adaptation purposes is insignificant. There are many methods such as channel sounding techniques, and estimation prediction methods that can be applied to further reduce the transmission of CSI in the feedback channel. In our analysis of the throughput of information we hence ignore the transmission in the feedback channel.

This paper deals with channel coding only. One should keep in mind however that source and channel coding must be coordinated in a systematic fashion in order to achieve optimum performance [13], [10], [14]. Moreover, without attempting to deal with the
source coding problem, we would like to draw the readers' attention to the fact that the scheme presented in this paper can be used as an adaptive source/channel coding scheme where the family of pragmatic codes can be employed, for instance, to provide needed unequal protection for transmission of speech.

The paper is organized as follows: In section II, the pragmatic approach to trellis-coded modulation is discussed, a suitable mapping scheme is proposed, and its applicability to adaptive systems is explained. In section III, we introduce the adaptive scheme and the various elements needed for its design including a suitable approach to interleaving/deinterleaving. Section IV contains a detailed analysis of the BER and throughput performance of ATCMPSK in Rayleigh fading channels. The theoretical analysis of error performance are based on the classical transfer function approach. Finally, in Section V, simulation results are reported.

**II. PRAGMATIC TCM**

Viterbi et al. [12] proposed a pragmatic approach to trellis-coded modulation which is based on the realization of rate \( \frac{n}{n+1} \) trellis-coded schemes using a single rate 1/2 encoder/decoder in conjunction with an MPSK signal mapper where \( M = 2^{n+1} \). The name pragmatic refers to code implementation employing the widely used, best known rate 1/2 convolutional encoder, and the corresponding Viterbi decoder. Figure 1 shows the pragmatic encoder, and the corresponding trellis diagram for the "good" convolutional code of constraint length \( K = 3 \).

The mapping scheme suggested in [12] can be called *sectorized-Gray-coded mapping*, and it works as follows: the \( (n+1) \) bits at the output of the encoder are used to select a signal in the signal space. The two least significant bits, which are the output of the rate 1/2 encoder, select one out of four signals according to Gray code. The remaining \( (n-1) \) bits choose a sector in the signal space lexicographically, as shown in Figure 2 for the

*Good codes refer to codes optimized for minimum Hamming distance.*
case of 16PSK. Note that equivalent signals in different sectors (signals with identical two least significant bits) belong to parallel branches of the same state transition.

In this paper, we suggest a different mapping called double-Gray-coded mapping. As for sectorized-Gray-coded mapping, the two least significant bits corresponding to the output of the rate 1/2 encoder select one out of four signals according to Gray code. The remaining (n-1) bits choose a sector in the signal space again according to Gray code, as depicted in Figure 2 for 16PSK signalling scheme. The first two bits of the encoded symbol determine the sector to which the signal belongs. Note that both mapping schemes result in the same signal space for constellations with less than 16 signals.

The obvious advantage of the double-Gray-coded mapping scheme for 16PSK is that signals belonging to parallel branches of neighbouring sectors differ in one bit as compared to 2 bits for sectorized-Gray-coded mapping. This improves the error performance of the code since most errors happen in neighbouring sectors. The BER performance of the trellis-coded 16PSK signals using the two different mapping schemes are compared in [15]. The mapping scheme employed throughout this research is double-Gray-coded MPSK, unless otherwise stated.

It is shown in [12] that the asymptotic coding gains (ACG) of the pragmatic codes in the AWGN channel are almost identical to Ungerboeck's codes. Pragmatic code, however, are much simpler to implement, and the same decoder can be used for different coding/modulation (codulation) rates.

Although pragmatic trellis codes perform nicely in the AWGN environment, they are not nearly as good in the fading environment. The major reason for this degradation in performance is the presence of parallel transitions in their state diagram which gives rise to single signal error events. Consequently, for good BER performance, lower rate pragmatic codes must be used—since going into higher rates translates to poor performance during deep fades. A fixed low rate code, however, sacrifices the throughput even during good channel conditions. Therefore, to take advantage of periods of good
channel conditions, a system is needed that can increase its rate of transmission during those periods.

Adaptive schemes have shown to be an effective method for combating the time varying nature of the distortion in the fading channel. A series of published papers report considerable gains by adapting certain parameters in the transmitter according to channel conditions—from symbol duration and power to transmission rate [9], [16] – [18].

III. SYSTEM MODEL

The model for the fading channel used in this study is that presented by Jakes in [19]. The envelope of the fading process is Rayleigh distributed with probability density function

\[ f(\alpha) = 2\alpha \exp\left(-\alpha^2\right) \quad \alpha \geq 0. \] (1)

The phase of the fading signal is uniformly distributed. We assume that the effect of phase on the received signal is fully compensated for by the receiver. This can be done in practice by pilot tone insertion and calibration [20], [21], or tracking the phase by a phase locked loop [22]. We should keep in mind, however, that in a fading environment, the exact tracking of the phase is extremely difficult. In fact, differential detection seems to be one of the only viable solutions for this problem. Simon and Divsalar [23] have devised a differential detection scheme for MPSK signals called *multiple-symbol differential detection*, which is based on maximum-likelihood sequence estimation. Their idea can be applied to differential detection of TCMPSK signals.

The effect of fading on the amplitude of the transmitted signal is the same as multiplying the in-phase and quadrature components of the baseband signal by the time varying fading process \( \alpha(t) \). Since the transmitted signal is attenuated by the fading component \( \alpha(t) \), the square of the Euclidean distance is no longer an optimum metric. The *Gaussian metric* which takes into account the effect of fading can easily be shown to
If $x_i$ is an arbitrary signal in the signal space, $y_i$ is the received signal, and $\alpha_i$ is the amplitude of the fading process at the time of transmission, then the corresponding Gaussian metric is [7]:

$$m_i = |y_i - \alpha_i x_i|^2.$$  \hspace{1cm} (2)

A. Adaptive Trellis-Coded Multiple-Phase-Shift-Keying (ATCMPSK)

A simplified block diagram of the proposed ATCMPSK scheme is shown in Figure 3. Information bits enter the information buffer at rate $R_b$ bits/sec. The buffer is needed since the effective rate of information is a function of the conditions in the time varying channel while the symbol transmission rate $R_s = \frac{1}{T}$ is constant, where $T$ denotes symbol duration. To choose the suitable scheme for transmission, we need channel state information (CSI) which in our model is the amplitude of fading $\alpha(kLT)$ at the start of the transmission of each L-bit block. CSI is supplied by the fading estimator to the transmitter for adaptation, and to the receiver for decoding purposes. As shown in the block diagram, the receiver is assumed to have ideal CSI at the start of each symbol transmission.

The transmitter decides what scheme should be used at the start of every transmission according to a given set of thresholds chosen to keep the BER below a certain level. There are five possible schemes to choose from:

1. Rate 1/2 TCQPSK with 3 repetitions
2. Rate 1/2 TCQPSK with 2 repetitions
3. Rate 1/2 TCQPSK
4. Rate 2/3 TC8PSK
5. Rate 3/4 TC16PSK.

Rate 1/2 TCQPSK with 3 repetitions is used during the worst, and rate 3/4 TC16PSK during the best channel conditions. Since there are five different codes employed by
the ATCMPSK under study, we have four adaptation thresholds: $\mu_1, \mu_2, \mu_3, \mu_4$. The transmitter looks at the value of the immediate SNR defined as: $SNR_i = \alpha_i^2 E_b / N_0$, and determines the suitable codulation scheme. If $0 \leq SNR_i \leq \mu_1$ rate 1/2 TCQPSK with 3 repetitions is employed. Otherwise, if $\mu_1 \leq SNR_i \leq \mu_2$ rate 1/2 TCQPSK with two repetitions is chosen, and so on. The number of schemes can be increased in order to attain larger throughput during very good channel conditions, and superior error performance during

Going back to the block diagram, information bits are fed into a rate 1/2 convolutional encoder, passed to the interleaver, and combined with the appropriate number of information bits (according to CSI) to form the adaptively encoded channel symbol. The encoded channel symbol is then mapped into a baseband signal according to double-Gray-coded mapping. The baseband signals are modulated by the carrier, and sent over the channel. The only sources of distortion are assumed to be the AWGN and the envelope attenuation due to fading. The demodulator recovers the distorted in-phase and quadrature (I/Q) components of the baseband signal which are then passed to the deinterleaver which reestablishes the correct order of rate 1/2 encoded symbols. However, as it will be discussed shortly, the order in which appended information bits are transmitted is not restored until the decoding function has been performed. Finally, the decoder, which is a slightly modified rate 1/2 Viterbi decoder, decodes the received sequence of symbols, and delivers them to the reorganizer which restores the original order of the information sequence. The Viterbi decoder is capable of constructing parallel branches in its trellis diagram. Changing to a different modulation scheme is then equivalent to changing the number of parallel branches in each state transition. The output of the reorganizer is an estimate of the information bits produced by the source. Note that the same encoder/decoder is used for different codulation rates.

---

* Adaptation thresholds are chosen according to the performance of the individual schemes in AWGN.
The MPSK transmitter can accommodate repetition codes. The concept of repetition codes is based on repeating the information to lower the BER [24]. During bad channel conditions, the signal representation of the encoded symbol is transmitted \( R \) times. This has the same effect as multiplying the signal energy by \( R \), and may also help the signal recover from a relatively short fade by increasing signal duration. Of course, the cost of improvement in BER performance is a reduction in throughput.

To accommodate repetition codes, the decoder should allow for multiple signals on the same branch. That is, the metrics for the received repeated signals are added up for the calculation of the final metric. For example, if \( x_i \) is an arbitrary signal in the signal space, \( R \) repetitions have been applied, \( Y_i = (y_{i1}, y_{i2}, ..., y_{iR}) \) is the received sequence of repeated signals, and \( A_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{iR}) \) are the corresponding fading amplitudes, then the branch metric for the transition corresponding to branch signal \( x_i \) is:

\[
m_i = \sum_{j=1}^{R} |y_{ij} - \alpha_{ij}x_i|^2.
\]  

(3)

B. Interleaving/Deinterleaving

Interleaving is traditionally performed to convert channels with memory to memoryless channels. It effectively removes the correlation between the samples of Rayleigh faded signal, and hence combats the disastrous effect of deep fades on the reconstruction procedure in the trellis during decoding [25]. Conventional interleaving and deinterleaving consists of storing encoded symbols into a matrix in successive rows, and transmitting the symbols in successive columns thus removing the correlation between the symbols in adjacent branches of the trellis. The interleaving matrix \( Z_i \) has size \( d \times s \) where \( d \) is called the interleaving depth and \( s \) the interleaving span. The deinterleaver performs the opposite operation. It stores the received symbols in successive columns, and delivers them to the decoder in successive rows, and hence restores the original order of the transmitted symbols.
Conventional interleaving/deinterleaving is not appropriate for ATCMPSK. Storing the encoded symbols in a matrix introduces an unavoidable delay. By the time an interleaved symbol is sent for transmission, channel conditions may have changed, and the encoded signal may no longer have the suitable scheme for transmission. We should therefore devise a new interleaving method. Figure 4 shows the structure of the interleaver for the proposed scheme. Information bits are passed to the encoder, and the encoded symbols are then delivered to the interleaver where they are stored in a serial fashion in successive rows of the interleaving matrix $Z_i$ until it is filled. The commutator receives CSI from the channel estimator. If channel conditions allow for higher level codulation schemes (TC8PSK, TC16PSK), the commutator will redirect the information bit flow to the adaptive mapper, which will take in the appropriate number of information bit(s) (1 for 8PSK, or 2 for TC16PSK) together with rate 1/2 encoded symbols which are delivered from the interleaver in successive columns, to form the channel symbol. If channel conditions dictate the use of TCQPSK, the rate 1/2 encoded symbol in the interleaving matrix will be mapped into a QPSK signal (no uncoded bits). Should channel conditions require repetitions, the mapped QPSK signal is repeated $R$ times. This procedure is repeated, symbol by symbol, until all interleaved symbols are used.

The deinterleaving function is depicted in Figure 5. Received symbols are stored in the successive columns of the deinterleaving matrix $Z_d$, and are then delivered in successive rows to the Viterbi decoder. By this time, the original order of the rate 1/2 encoded symbols has been restored, but the appended bits are out of order, and cannot be delivered until the whole sequence of symbols (of length equal to the interleaving size) is decoded. The order of the sequence of information bits is then restored by the reorganizer which has the knowledge of the past history of received symbols.

IV. PERFORMANCE ANALYSIS
A. Upper Bound on BER

For the purpose of theoretical analysis, the following assumptions are made:

- The channel is frequency nonselective, and slowly fading.
- The effect of phase is fully compensated for by the receiver.
- Ideal pulse shaping, and hence no ISI.
- Ideal channel state information (CSI) is available at the receiver side.
- Symbol by symbol adaptation is applied.
- We have ideal interleaving defined as follows:
  1. All rate 1/2 encoded symbols are faded by uncorrelated fading samples.
  2. Fading component is kept constant for repeated versions of the same transmitted signal.

We have assumed ideal symbol by symbol adaptation. That is, the transmitter adjusts its rate according to ideal CSI obtained from the fading estimator at the start of every symbol transmission interval, and the decoder adjust its parameters at every decoding instant. The metric at every decoding instant is calculated according to the CSI, and is thus a function of the fading amplitude.

Let $X_L = (x_1, x_2, \ldots, x_L)$, $X'_L = (x'_1, x'_2, \ldots, x'_L)$ be two sequences of encoded symbols with length $L$, and $P(X_L \rightarrow X'_L)$ be the pairwise error probability; that is, the probability of choosing $X'_L$ given $X_L$ was transmitted. The union bound on the probability of an error event can then be expressed as [26]:

$$P(e) \leq \sum_{L=1}^{\infty} \sum_{X'_L \neq X_L} \sum_{X_L} P(X_L) P(X_L \rightarrow X'_L),$$

or equivalently:

$$P(e) \leq \sum_{L=1}^{\infty} \sum_{X_L} P(X_L) \sum_{X'_L \neq X_L} E[P(X_L \rightarrow X'_L|x_1, x_2, \ldots, x_L)].$$
Assuming maximum-likelihood detection [27, p. 247]:

\[ P(X_L \rightarrow X'_L | \alpha_1, \alpha_2, ..., \alpha_n) = Q \left( g_{M(n)}(X_L, X'_L) \sqrt{\frac{E_s}{2N_0}} \right) \quad (6) \]

where

\[ g_{M(n)}^2(X_L, X'_L) = \sum_{n=1}^{L} \alpha_n^2 R_n \| f_{M(n)}(x_n) - f_{M(n)}(x'_n) \|^2 \quad (7) \]

is the Gaussian distance between the two sequences. Note that \( \| \cdot \| \) represents Euclidean distance, and \( f_{M(n)}(\cdot) \) is the nonlinear mapping function for the MPSK signal space normalized on \( E_s \). where \( M(n) \) can take on 3 different values: 4, 8, and 16 for distances in QPSK, 8PSK, and 16PSK signal spaces, respectively. \( R_n \) is the number of repetitions employed in the signalling interval \( n \) which is either 1 (for QPSK, 8PSK, and 16PSK), or equal to the number of transmissions when repetitions are applied. Applying the bound

\[ Q(y) \leq \frac{1}{2} \exp \left( -\frac{y^2}{2} \right) \quad y \geq 0, \quad (8) \]

we get

\[ P(X_L \rightarrow X'_L | \alpha_1, \alpha_2, ..., \alpha_n) \leq \frac{1}{2} \exp \left( -\frac{E_s}{4N_0} g_{M(n)}^2(X_L, X'_L) \right) \]

\[ = \frac{1}{2} \exp \left( -\frac{E_s}{4N_0} \sum_{n=1}^{L} \alpha_n^2 R_n d_{M(n)}^2(x_n, x'_n) \right) \quad (9) \]

\[ = \frac{1}{2} \prod_{n=1}^{L} \exp \left( -\frac{R_n E_s}{4N_0} \alpha_n^2 d_{M(n)}^2(x_n, x'_n) \right), \]

where

\[ d_{M(n)}^2(x_n, x'_n) = \| f_{M(n)}(x_n) - f_{M(n)}(x'_n) \|^2 \quad (10) \]

is the euclidean distance between the two signals. Averaging over the fading intervals we then get:

\[ E[P(X_L \rightarrow X'_L | \alpha_1, \alpha_2, ..., \alpha_L)] = \frac{1}{2} \prod_{n=1}^{L} \int f(\alpha_n) d\alpha_n. \quad (11) \]
The scheme employed at each signalling interval depends on the amplitude of fading in that interval. The BER performance of a scheme with \( R_n \) repetitions and energy \( E_s \) is the same as its performance without any repetitions and energy \( R_n E_s \). Moreover, when repetition schemes are applied, it is assumed that all \( R_n \) symbols are equally attenuated. For all other schemes, fading is assumed constant over one symbol period only.

The adaptation thresholds for immediate SNR are: \( \mu_1, \mu_2, \mu_3, \mu_4 \). We can determine the adaptation thresholds in terms of the fading amplitude in the following way:

\[
\mu_i \leq \alpha_n^2 \frac{E_s}{N_0} \leq \mu_{i+1} \Rightarrow \sqrt{\frac{\mu_i}{E_s/N_0}} \leq \alpha_n \leq \sqrt{\frac{\mu_{i+1}}{E_s/N_0}} .
\]  

(12)

The adaptation thresholds for the fading amplitude are then determined by the relationship \( \sqrt{\frac{\mu_i}{E_s/N_0}} = \nu_i \). Therefore, the integral in (11) becomes:

\[
\int_{\nu_1}^{\nu_2} \exp \left( \frac{-3E_s}{4N_0} \alpha_n^2 d_4(x_n, x'_n) \right) f(\alpha_n) d\alpha_n + \int_{\nu_2}^{\nu_3} \exp \left( \frac{-2E_s}{4N_0} \alpha_n^2 d_4(x_n, x'_n) \right) f(\alpha_n) d\alpha_n
\]

\[
+ \int_{\nu_3}^{\nu_4} \exp \left( \frac{-E_s}{4N_0} \alpha_n^2 d_4(x_n, x'_n) \right) f(\alpha_n) d\alpha_n + \int_{\nu_4}^{\nu_5} \exp \left( \frac{-E_s}{4N_0} \alpha_n^2 d_4(x_n, x'_n) \right) f(\alpha_n) d\alpha_n
\]

\[
+ \int_{\nu_5}^{\nu_6} \exp \left( \frac{-E_s}{4N_0} \alpha_n^2 d_4(x_n, x'_n) \right) f(\alpha_n) d\alpha_n ,
\]  

(13)

where \( f(\alpha_n) \) is the Rayleigh density function. Furthermore,

\[
\int_{\nu_i-1}^{\nu_i} \exp \left( \frac{-R_i \alpha^2 E_s d_4^2(x_n, x'_n)}{4N_0} \right) 2\alpha \exp (-\alpha^2) d\alpha
\]

\[
= \frac{\exp (-\nu_i^2 \beta_{n,i} E_s/N_0) - \exp (-\nu_{i-1}^2 \beta_{n,i} E_s/N_0)}{\beta_{n,i} E_s/N_0}
\]

\[
= \frac{\exp (-\mu_{i-1} \beta_{n,i}) - \exp (-\mu_i \beta_{n,i})}{\beta_{n,i} E_s/N_0} ,
\]

(14)

where

\[
\beta_{n,i} = \frac{1}{E_s/N_0} + \frac{R_i d_4^2(x_n, x'_n)}{4} .
\]

(15)
Substituting (14) into (5) we obtain:

\[
P(e) \leq \frac{1}{2} \sum_{L=1}^{\infty} \sum_{X_L} P(X_L) \sum_{X'_{L} \neq X_L}^{L} \prod_{n=1}^{L} \left( \frac{1 - \exp(-\mu_1 \beta_n,1)}{\beta_n,1 E_s/N_0} + \frac{\exp(-\mu_2 \beta_n,2) - \exp(-\mu_2 \beta_n,2)}{\beta_n,2 E_s/N_0} + \frac{\exp(-\mu_3 \beta_n,3) - \exp(-\mu_3 \beta_n,3)}{\beta_n,3 E_s/N_0} + \frac{\exp(-\mu_4 \beta_n,4) - \exp(-\mu_4 \beta_n,4)}{\beta_n,4 E_s/N_0} + \frac{\exp(-\mu_5 \beta_n,5)}{\beta_n,5 E_s/N_0} \right).
\]

(16)

Denoting the term inside the brackets in (16) by \( h(x_n, x'_n) \), the upper bound on the error-event probability can be expressed as follows:

\[
P(e) \leq \frac{1}{2} \sum_{L=1}^{\infty} \sum_{X_L} P(X_L) \sum_{X'_{L} \neq X_L}^{L} \prod_{n=1}^{L} h(x_n, x'_n).
\]

(17)

The right hand side of (17) is simply the transfer function of the error-state diagram of the code. For pragmatic trellis codes, this transfer function is a scalar value [26]. As an example, the error-state diagrams for pragmatic codes of constraint length 3 have the general structure shown in Figure 6. The labels of the error-state diagram are the \( h(x_n, x'_n) \) terms corresponding to each transition on the error state diagram. We must multiply each term by \( 1^k \), where \( k \) is the number of bits in error corresponding to that term. The transfer function of the above error-state diagram is then found to be [26]:

\[
T\left( \frac{E_s}{N_0}, I \right) = \frac{1}{1 - g_4 g_5} \left[ \frac{g_1 g_2 g_3 g_7}{(1 - g_4 g_5)(1 - g_6) - g_2 g_3 g_5} + g_1 g_4 g_7 \right].
\]

(18)

The upper bound on BER can be expressed by:

\[
P_b \leq \frac{1}{2n} \left| \frac{\partial T \left( \frac{E_s}{N_0}, I \right)}{\partial I} \right|_{I=1},
\]

(19)

where \( \bar{n} \) is the average number of information bits per adaptation interval, and \( P_\parallel \) is the probability of error due to parallel transitions. \( \bar{n} \) can be found as follows:
\[ \pi = P(TCQPSK) + 2P(TC8PSK) + 3P(TC16PSK) \]
\[= \int_0^{\mu_2} 2\alpha \exp(-\alpha^2) \, d\alpha + \int_{\mu_3}^{\mu_4} 2\alpha \exp(-\alpha^2) \, d\alpha + 3 \int_{\mu_4}^{\infty} 2\alpha \exp(-\alpha^2) \, d\alpha \]
\[= 1 + \exp\left(-\frac{\mu_3}{E_s/N_0}\right) + \exp\left(-\frac{\mu_4}{E_s/N_0}\right). \quad (20) \]

\(P_{\parallel}\) is not accounted for through the generating function approach. To find \(P_{\parallel}\), we must note that parallel transitions contribute to the probability of error only when channel conditions dictate the use of TC8PSK and TC16PSK. Therefore:
\[ P_{\parallel} = \int_{\mu_3}^{\mu_4} P_{\parallel, 8PSK} 2\alpha e^{-\alpha^2} \, d\alpha + \int_{\mu_4}^{\infty} P_{\parallel, 16PSK} 2\alpha e^{-\alpha^2} \, d\alpha. \]
\[= \int_{\mu_3}^{\mu_4} \frac{1}{2} Q\left(\alpha \sqrt{\frac{2E_s}{N_0}}\right) 2\alpha e^{-\alpha^2} \, d\alpha + \int_{\mu_4}^{\infty} \frac{2}{3} Q\left(\alpha \sqrt{\frac{E_s}{N_0}}\right) 2\alpha e^{-\alpha^2} \, d\alpha, \quad (21) \]
where
\[ P_{\parallel, 8PSK} = \frac{1}{2} Q\left(\alpha \sqrt{\frac{2E_s}{N_0}}\right) \quad (22) \]
is the probability of error due to parallel transitions when 8PSK is employed, and
\[ P_{\parallel, 16PSK} = \frac{2}{3} Q\left(\alpha \sqrt{\frac{E_s}{N_0}}\right) \quad (23) \]
is the probability of error due to parallel transitions when 16PSK is employed \([15]\).
Therefore:
\[ P_{\parallel} = \int_{\mu_3}^{\mu_4} \frac{1}{2} Q\left(\alpha \sqrt{\frac{2E_s}{N_0}}\right) 2\alpha e^{-\alpha^2} \, d\alpha + \int_{\mu_4}^{\infty} \frac{2}{3} Q\left(\alpha \sqrt{\frac{E_s}{N_0}}\right) 2\alpha e^{-\alpha^2} \, d\alpha. \]
\[\quad (24) \]
Applying (8) and integrating the above expression, we find that:
\[ P_{\parallel} \leq \frac{1}{4} e^{-\mu_3 \left(1 + \frac{2}{E_s/N_0}\right)} - \frac{\mu_3 \left(1 + \frac{2}{E_s/N_0}\right)}{1 + \frac{E_s}{2N_0}} \]
\[+ \frac{1}{3} e^{-\mu_4 \left(1 + \frac{1}{E_s/N_0}\right)} \quad (25) \]
\[+ \frac{1}{3} e^{-\mu_4 \left(1 + \frac{1}{E_s/N_0}\right)} \quad (25) \]
B. Throughput of ATCMPSK

Due to its adaptive nature, the throughput of the ATCMPSK scheme varies as a function of the immediate SNR. The throughput can be defined as the average number of information bits transmitted per symbol duration $T_s$.

The number of information bits per channel symbol is $1/3$ for QPSK with three repetitions, $1/2$ for QPSK with two repetitions, one for TCQPSK, two for TC8PSK, and three for TC16PSK. Therefore:

$$\bar{\eta} = \frac{1}{3}P(0 \leq \alpha \leq \nu_1) + \frac{1}{2}P(\nu_1 \leq \alpha \leq \nu_2) + P(\nu_2 \leq \alpha \leq \nu_3)$$
$$+ 2P(\nu_3 \leq \alpha \leq \nu_4) + 3P(\nu_4 \leq \alpha \leq \infty).$$

(26)

Since samples of fading are Rayleigh distributed, we have:

$$P(\nu_i \leq \alpha \leq \nu_{i+1}) = \exp(-\nu_i^2) - \exp(-\nu_{i+1}^2).$$

(27)

Recall that

$$\nu_i = \sqrt{\frac{\mu_i}{E_s/N_0}}.$$ 

(28)

The throughput of the proposed ATCMPSK can hence be expressed as:

$$\bar{\eta} = \frac{1}{3} + \frac{1}{6} \exp\left(-\frac{\mu_1}{E_s/N_0}\right) + \frac{1}{2} \exp\left(-\frac{\mu_2}{E_s/N_0}\right)$$
$$+ \exp\left(-\frac{\mu_3}{E_s/N_0}\right) + \exp\left(-\frac{\mu_4}{E_s/N_0}\right).$$

(29)

In general, for a system employing up to $R$ repetitions and $2^{n+1}$PSK, the throughput can be expressed as:

$$\bar{\eta} = \frac{1}{R} + \sum_{i=0}^{R-2} \frac{\exp\left(-\frac{\mu_{i+1}}{E_s/N_0}\right)}{(R-i)(R-i-1)} + \sum_{i=0}^{n-2} \exp\left(-\frac{\mu_{R+i}}{E_s/N_0}\right).$$

(30)
The BER, and throughput performance curves obtained by simulations and theory are given in Figure 7. To simulate ideal interleaving, we have used totally uncorrelated samples with a Rayleigh fading distribution. During repetition periods signal attenuation due to fading was kept constant. Simulation results correspond to that of theory which, at high SNR, is approximately 2 dB away from simulation results. On the same graph, we have plotted throughput curves obtained by theory and simulations which overlap since our derivation of throughput is exact.

V. SIMULATION RESULTS

A. The effect of Error Thresholds

The adaptation thresholds are chosen—using the performance curves of individual schemes in AWGN—to keep the BER below an error roof. Figure 8 depicts this concept. As the error roof is lowered, the thresholds for all schemes move to the right. That is, schemes with better BER performance (and lower throughput) are employed more often. At some point, slopes of the performance curves approach infinity, and, therefore, further lowering of the roof does not have any effect on the thresholds. Figure 9 shows the BER and throughput performance of ideally interleaved ATCMPSK with different error roofs obtained by simulations.

As expected, the throughput of the schemes using larger error roofs is significantly better than those with smaller ones. Therefore, the question: "what are the optimum thresholds?" is not easy to answer. In fact, because of the trade-off between error performance and throughput, it is impossible to quantify the difference between two schemes using different error roofs. One can observe the BER and throughput curves at a certain SNR and decide which one is more suitable for a certain application. Note that at high SNR, the difference between throughputs becomes small while the BER
performances are significantly different. Therefore, if the system is to operate at high SNR, it is better to use a relatively small error roof.

Introducing error roofs is not the only possible way of choosing the thresholds. In fact, we can obtain an infinite family of performance and throughput curves within the constraints that bound the BER and the throughput performance.

B. The effect of BT Product and Interleaving

Since we are considering only the effect of fading amplitude on the transmitted signal, increasing the BT product (B is the maximum Doppler frequency and T the symbol duration) will simply result in less correlated adjacent signals. Therefore, the performance of the system improves as BT is increased.* Figure 10 shows the performance curves for the adaptive scheme for different BT products. On the same graph, we have the curves for ideally interleaved ATCMPSK, and finite interleaving size of 600 \times 100. The performance of the ideally interleaved case at low SNR is shown to be slightly worse than the finite interleaved case. This is due to the nature of simulations. For ideal interleaving, all samples of fading are uncorrelated, except during repetitions where repeated signals are faded equally. For finite interleaving, however, repeated signals are faded according to channel conditions at the time of their transmission.

Figure 11 depicts the effect of finite interleaving. As expected, the BER performance improves with increasing interleaving size until no further improvement is possible.

C. The effect of Block Adaptation

So far, all simulations have assumed symbol by symbol adaptation. The effect of adapting the codulation rate to channel conditions at fixed intervals greater than one symbol duration is given in Figure 12.

* One must be aware, however, that in most practical applications, the BER performance becomes progressively worse as the BT product is increased. The degradation in performance is reflected through error floors which occur at higher error probabilities for faster fading.
As expected, the BER performance degrades as the duration of the adaptation interval is increased. This is due to the fact that channel conditions at the start of an adaptation interval may change significantly before the next interval is started. Moreover, the effect of the adaptation interval is a function of the fading process. For fast fading, a few symbol durations translate to considerable change in channel conditions, and an adaptation interval of such length will therefore degrade the performance significantly. However, in slow fading, as shown in Figure 12, the adaptation interval can be increased without paying a big price in performance.

Note that the effect of a constant delay in the feedback channel is exactly the same as block adaptation. Any degradation in the BER performance due to the delay in the feedback channel is similarly the result of an inaccurate CSI at the time of transmission.

D. Comparison with Fixed Rate Pragmatic TCM Schemes

Figure 13 compares the performance of ideally interleaved ATCMPSK schemes with fixed rate pragmatic trellis codes in Rayleigh fading channel. As shown, there is a coding gain of 3 to 20 dB depending on the SNR value. On the same graph, it is shown that ATCMPSK under fading is about 4dB away from the performance of fixed rate pragmatic trellis codes in AWGN*. In fact, using more diverse TCM schemes of different rates and more repetition in ATCMPSK, and hence more complexity, we can get even closer to the performance of best fixed rate codes in AWGN.

Fixed rate trellis codes with better BER performance (no parallel transitions) can be found [28], but the resulting coding gain is in no way comparable to that achieved by ATCMPSK.

* The performance of 4-state pragmatic trellis codes is practically the same as 4-state Ungerboeck trellis codes.
VI. CONCLUSIONS

Although fixed rate trellis codes are comparably good in terms of BER performance in Rayleigh fading, they fail to explore the time varying nature of these channels, and are often designed conservatively to combat very poor channel conditions. The fixed rate of transmission hence results in sacrificing the effective information rate during good channel conditions.

We have proposed an adaptive trellis-coded scheme called ATCMPSK which explores the time varying nature of Rayleigh fading channels. The ATCMPSK is extremely effective in combating fading, and is pragmatic in terms of its design and implementation, and results in 3–20 dB gain in BER performance when compared to fixed rate trellis codes. These gains are obtained using a simple encoder of constraint length three, under ideal conditions. We have investigated the loss in performance due to some nonideal conditions, and shown that even under those conditions one can achieve a very substantial gain using our adaptive scheme.

A suitable method to remove the memory in the channel has been introduced. It has been shown that there is lot to be gained from interleaving specially at high SNR values where high rate schemes are prevalent.

We have suggested that one way of determining the adaptation thresholds is to use the BER performance curves (in the AWGN channel) of the pragmatic schemes used in ATCMPSK. However, there are many other ways in which the thresholds can be chosen. We can in fact shape the BER curves, within the performance constraints of the scheme, by choosing different sets of thresholds.

We have used the conventional transfer function approach in order to find an upper bound for the BER performance, and used the statistics of the Rayleigh fading channel to formulate an exact expression for the throughput. Our theoretical results are consistent in terms of their agreement with simulations.
We have only investigated the performance of our scheme in Rayleigh fading channels. It is clear however that the same scheme can be employed in a variety of channels such as fading mobile satellite channels characterized by Rician Fading.
Figure 1 Pragmatic encoder and its trellis structure for $K=3$. 

a) Pragmatic Encoder

$K=3$

c) 8PSK

d) 16PSK

b) QPSK
Figure 2 Double-Gray-coded mapping for 16PSK.

Figure 3 Simplified block diagram of ATCMPSK.
Figure 4 Interleaver Structure for ATCMPSK.

Figure 5 The structure of the deinterleaver for ATCMPSK.

Figure 6 The general structure of the error-state diagram for pragmatic codes (K=3).
Figure 7  Theoretical and simulation curves for BER and throughput performance of ATCMPSK under Rayleigh Fading.
Figure 8 Adaptation thresholds for ATCMPSK using different error roofs.
Figure 9 BER and throughput performance of ideally interleaved ATCMPSK with different error thresholds.
ATCMPSK, $K=3$, Rayleigh Fading, Error Roof = 0.01

Delayless Feedback, Symbol Adaptation

Finite Interleaving
(600 x 100) BT = 10^{-5}

Ideally Interleaved

Figure 10 The effect of BT product on the BER performance of ATCMPSK
Figure 11 The effect of finite interleaving on the BER performance of ATCMPSK.
Figure 12 The effect of block adaptation on the BER performance of ATCMPSK.

ATCMPSK, $K=3$
Rayleigh Fading
$BT = 10^{-5}$
No Interleaving
Delayless Feedback

Block Len. = 200

Block Len. = 100

Symbol by Symbol

Error Roof = 0.01
Block Len. = 50
Figure 13 Throughput versus SNR curves for ideally interleaved ATCMPSK, and ideally interleaved fixed rate pragmatic codes in Rayleigh fading and AWGN channels.
Bibliography


