Adaptive Bit-Interleaved Coded Modulation

Pınar Örmeci, Xueting Liu, Dennis L. Goeckel, and Richard D. Wesel, *Member, IEEE*

***Abstract*—Adaptive coded modulation is a powerful method for achieving a high spectral efficiency over fading channels. Recently proposed adaptive schemes have employed set-partitioned trellis-coded modulation (TCM) and have adapted the number of uncoded bits on a given symbol based on the corresponding channel estimate. However, these adaptive TCM schemes will not perform well in systems where channel estimates are unreliable, since uncoded bits are not protected from unexpected fading. In this paper, adaptive bit-interleaved coded modulation (BICM) is introduced. Adaptive BICM schemes remove the need for parallel branches in the trellis—even when adapting the constellation size, thus making these schemes robust to errors made in the estimation of the current channel fading value. This motivates the design of adaptive BICM schemes, which will lead to adaptive systems that can support users with higher mobility than those considered in previous work. In such systems, numerical results demonstrate that the proposed schemes achieve a moderate bandwidth efficiency gain over previously proposed adaptive schemes and conventional (nonadaptive) schemes of similar complexity.**

***Index Terms*—Bit-interleaved coded modulation, time-varying fading channels, trellis-coded modulation.**

I. Introduction

THE VISION of a communication system that provides ubiquitous high-speed access to information has led to widespread research in recent years on information transmission over multipath fading channels. One possible method to achieve robust and spectrally efficient communication over multipath fading channels is to adapt the transmission scheme to the current channel characteristics using channel estimates available at the transmitter. Unlike nonadaptive schemes, which are designed for worst-case channel conditions to achieve acceptable performance, adaptive signaling methods take advantage of favorable channel conditions by allocating power and rate efficiently.

The promise of a particular communications architecture can generally be analyzed through information theory by determining the channel capacity of the proposed scheme. However, in the field of adaptive signaling, the results of information theory must be interpreted carefully. In particular, [1] indicates that for a system with independent and identically distributed (i.i.d.) Rayleigh fading affecting the coded symbols, the Shannon capacity with knowledge of the channel fading values at the transmitter and receiver is only negligibly larger than that when the current channel fading values are known only at the receiver. With the performance of systems employing iterative decoding of concatenated codes approaching channel capacities [2], this suggests that knowledge of the current channel fading values at the transmitter is of limited utility in systems that are not delay-constrained and permit high-complexity decoding. However, recent work [3]–[5] has shown that significant gains in bandwidth efficiency are exhibited by adaptive trellis-coded modulation (TCM) [6] schemes over their nonadaptive counterparts for delay-constrained, low-complexity decoders. Hence, while adaptive techniques may not significantly increase capacity, they may provide better performance for low-complexity or delay-constrained systems.

The proposed adaptive TCM schemes approach capacity with only a standard Viterbi decoder at the receiver [3] and can exhibit very small delay, requiring *no* interleaving at the transmitter if the channel is known perfectly and only limited interleaving if the channel estimates are outdated or noisy [7]. This makes adaptive schemes ideal for real-time applications where receiver complexity should be minimized. Recent results in information theory that capture the implications of limited delay have supported this last statement. Caire *et al.* [8] have demonstrated that, with respect to information outage or delay-limited capacity, knowledge at the transmitter of the entire frame of fading values that a codeword will experience can be employed to greatly improve system performance.

In recent years, numerous approaches have been proposed to optimize the transmitted signal using knowledge of the current channel fading value of a frequency-nonselective Rayleigh fading channel [1], [3]–[5], [9]–[12]. If the current fading value can be predicted perfectly at the transmitter, one generally modifies the transmitted signal constellation to approximately maintain a constant Euclidean distance between signal points of minimal separation after the fading value is applied to the transmitted signal [3]. Then, the only impairment to the perfect operation of the system is the additive noise, and constructions typical of additive white Gaussian noise (AWGN) channels are appropriate. Thus, these schemes have employed set-partitioned TCM and have adapted the number of uncoded bits on a given symbol based on the corresponding channel estimate. However, as noted in [4], [5], [13], when the channel estimate is outdated, the distribution of the current channel fading conditioned on the estimate has a Rician distribution. Furthermore, recent work by Goeckel [4], [5] has shown that, as the mobility of system users increases, the effective channel conditioned on outdated

Paper approved for publication by A. Goldsmith, the Editor for Wireless Communication of the IEEE Communications Society. Manuscript received December 20, 1999; revised October 16, 2000. This work was supported by the National Science Foundation under Grants NCR-9714597, CCR-9733089, and CCR-9875482. This paper was presented in part at the Thirty-Third Annual Conference on Information Sciences and Systems, Baltimore, MD, March 1999.

P. Örmeci and D. L. Goeckel are with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003 USA (e-mail: ormeci@ecs.umass.edu; goeckel@ecs.umass.edu).

X. Liu and R. D. Wesel are with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: xueting@ee.ucla.edu; wesel@ee.ucla.edu).

Publisher Item Identifier S 0090-6778(01)08166-1.

fading estimates becomes more Rayleigh, thus greatly impeding the ability to use uncoded bits for rate adaptation. In this case, adaptive TCM schemes, which employ uncoded bits, fail to be an efficient method of coded modulation. This is because error events from poor predictions dominate the bit error probability in an adaptive system in the same way that error events from bad fades dominate the bit error probability of nonadaptive wireless communication systems.

A solution for extending the applicability of adaptive signaling schemes to systems with higher mobility than previously considered is to employ bit-interleaved coded modulation (BICM), which has recently been considered for nonadaptive systems operating over the Rayleigh fading channel [14], [15]. In the nonadaptive case, BICM increases the time diversity of systems operating over fading channels by providing an independent fading component for each channel bit out of the convolutional encoder, as opposed to each channel symbol as in standard symbol-interleaved TCM systems [16], [17]. However, the BICM structure has even greater potential in adaptive coding, as it does not require the use of uncoded bits to adapt constellation size. The uncoded bits in adaptive TCM only achieve a diversity of one, but the inherent diversity of adaptive BICM protects every information bit from channel prediction errors.

To consider the robustness to prediction errors of the two structures mathematically, the impact of a single bad prediction on each scheme is considered. By analyzing the bit error rate (BER) performances of BICM and TCM under a single prediction error, it is shown that BICM is more robust to prediction errors than TCM. This motivates the employment of adaptive BICM to extend the user mobility range for which adaptive signaling will be effective.

Two types of constructions will be considered for the design of adaptive BICM. The first exploits the inherent robustness to poor predictions of the BICM architecture, while the second explicitly accounts for the anticipated prediction error. In the first construction, which will be termed the "deterministic design," the adaptive system chooses a modulation scheme under the assumption that the current channel fading value is predicted perfectly. The performance of the scheme is then characterized; of course, due to prediction errors, the performance goals of the system will not be met. An energy margin, which reduces the BER of the system at the expense of rate, is then incorporated into the system until the performance goal is met. Unlike adaptive TCM [4], [5], this design works effectively. It is the use of uncoded bits for rate adaptation that reduces the effectiveness of the deterministic design method in the adaptive TCM case. The second construction explicitly takes into account the variation of the fading channel between channel estimation and data transmission and is designed following the arguments given in [4], [5]. Bandwidth efficiencies similar to those for the first construction are obtained. Thus, unlike in [4] and [5], it is concluded that the selection of the current signal set based on models for the channel variation is not a key consideration; instead, using the characteristics of the model to recognize that BICM should be employed to correct for prediction errors under such variation is the critical step.

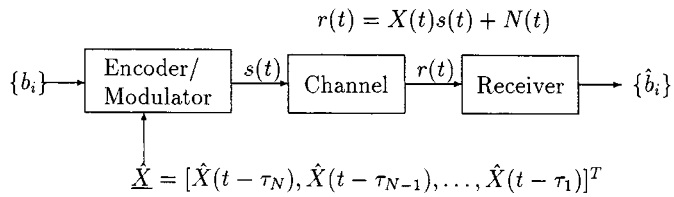

Fig. 1. System model.

Section II reviews the basics of adaptive signaling over time-varying channels. Section III analyzes the robustness of TCM and BICM structures to prediction errors. Section IV designs strongly robust (as defined in [4], [5]) adaptive BICM schemes for the time-varying channel. Section V gives numerical results and analysis for several adaptive BICM designs for fading channels. Finally, Section VI presents the conclusions.

II. System Model and Characterization

A. System Model

The baseband system model, which is identical to that of [4], [5], is illustrated in Fig. 1. An independent and identically distributed sequence of information bits $\{b_i\}$, each equally likely to be 0 or 1, is assumed. The transmitted signal is given by $s(t) = \sum_{k=-\infty}^{\infty} z_k p(t - kT_s)$, where z_k is the kth complex data symbol, $p(t)$ is a pulse shape that results in no intersymbol interference (ISI) in the samples (spaced at T_s) of the output of the matched filter at the receiver, and $1/T_s$ is the symbol rate. The additive noise $n(t)$ is white Gaussian noise with two-sided power spectral density $N_0/2$. A frequency-nonselective channel is considered in this work; henceforth, the fading is modeled as a complex multiplier $X(t)$. The Gaussian wide-sense stationary uncorrelated scattering fading model [18] is assumed, where the independent in-phase and quadrature components of $X(t)$ are zero mean Gaussian processes with autocorrelation function $R_X(\tau)$. Hence, the channel is modeled as a Rayleigh fading channel. It is also assumed that $X(t)$ varies slowly enough so that it can be considered as constant over a single symbol period. Per Fig. 1, the vector of outdated channel fading estimates, $\underline{\hat{X}} = \underline{\hat{X}}_R + j\underline{\hat{X}}_I$, where $\tau_{i+1} > \tau_i, \forall\, i$, is assumed to be available at the transmitter. It will be assumed that the outdated fading estimates are noiseless; that is, $\hat{X}(kT_s - \tau_i) = X(kT_s - \tau_i), i = 1, \ldots, N$, and thus $\underline{\hat{X}} = \underline{X}$, although the consideration of noisy estimates is conceptually identical [7]. Throughout this work, coherent reception with perfect channel state information (CSI) *at the receiver* is assumed.

B. Characterization of the Channel Variation

Adaptive signaling attempts to optimize the signal set using perfect or imperfect knowledge of the current fading values at the transmitter. Suppose the choice of the signal set for the kth symbol is being considered at the transmitter. Let $Y = |X(kT_s)|$ be the amplitude of the fading that multiplies the kth transmitted symbol. Throughout this work, the number of outdated fading estimates is assumed to be one, although the case of multiple outdated estimates with known $R_X(\tau)$ is conceptually identical [4], [5]. It is shown in [4], [5], and [13] that Y

is Rician when conditioned on $\underline{X}$ with conditional probability density function given by

$$p_{Y|\underline{X}}(y|\underline{x}) = \frac{y}{\sigma_Y^2} \exp\left(-\frac{y^2+s^2}{2\sigma_Y^2}\right) I_0\left(\frac{ys}{\sigma_Y^2}\right), \qquad y \geq 0 \tag{1}$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function. With one outdated estimate, normalizing such that $E[(X_R(kT_s))^2] = E[(X_I(kT_s))^2] = 1/2$ so that $E[|X(kT_s)|^2] = 1$ and defining the correlation coefficient $\rho = 2R_X(\tau_1)$, one obtains $s^2 = (|x|^2\rho^2)/2$ and $\sigma_Y^2 = (1-\rho^2)/2$. Define the specular-to-diffuse component (Rician) factor of (1) by $K = (s^2/2\sigma_Y^2) = (|x|^2\rho^2)/(2(1-\rho^2))$. As the mobility of system users increases (i.e., ρ decreases), K drops rapidly for any given estimate. Thus, as system user mobility increases, there will be a point where a large value of K is highly unlikely and thus adaptive TCM schemes, which employ uncoded bits for rate adaptation, will fail to be an efficient method of coded modulation.

In wireless systems, $R_X(\tau)$ is generally not known at the transmitter. Thus, an adaptive scheme should be able to operate below the desired bit error probability for all autocorrelation functions $R_X(\tau)$ in some class $\mathcal{R}$, where the class $\mathcal{R}$ is determined from propagation measurements. For the case of one outdated estimate, a class is characterized by $\rho_{\min} = \inf_{R_X(\tau)\in\mathcal{R}} R_X(\tau_1)$. In this paper, the adaptive methods will be designed to operate below the desired bit error rate for all $\rho \in [\rho_{\min}, 1]$.

III. Robustness to Prediction Errors

An exact mathematical comparison of the adaptive BICM and adaptive TCM structures, although possible, is complicated by the fact that the effective channel changes on a symbol-by-symbol basis, which in turn alters the modulation scheme employed. Thus, the resulting equations are cumbersome and yield little insight into the system operation. However, it was noted in simulations of the TCM scheme that for systems with relatively small values of $\rho_{\min}$, most errors are caused by a single bad prediction, which is analogous to a single bad fade causing errors for a conventional nonadaptive TCM scheme operating over a Rician fading channel. This section uses a modified transfer function bound to analyze the robustness of BICM and TCM schemes to a single prediction error. This analysis suggests that the code structure of BICM is more suitable for adaptive systems operating with relatively small values of $\rho_{\min}$.

A. Channel Model

With adaptive coding, if the channel is predicted perfectly at the transmitter, the minimum distance between constellation points after the fading is applied can be kept constant despite the time variation in the fading. As motivated above, suppose that the channel is estimated perfectly except during the transmission of the lth symbol, where a prediction error occurs. This implies that the minimum distance between constellation points remains constant except at the lth transmitted symbol period.

To model this behavior simply for application of a transfer function union bound, we consider a simpler scenario that captures the effect of a failed prediction without the complication of variation in the constellation size. We consider nonadaptive TCM and BICM on an AWGN channel (i.e., $|X(t)| = 1$) except that the lth symbol suffers a fade $Y = |X(lT_s)|$. This is analogous to an adaptive system with perfect channel estimation except for an estimation error on the lth symbol. The information bits encoded to produce the lth transmitted symbol suffer the most from this prediction error. Thus, our modified transfer function union bound estimates the BER only for those bits.

B. Transfer Function Bound on the Probability of Bit Error for TCM and BICM Under a Single Prediction Error in an AWGN Channel

For a trellis code, the standard transfer function is

$$T(W,I) = \sum_w \sum_i N(w,i) W^w I^i \tag{2}$$

where

- w Euclidean distance of an error event;
- i number of information bit errors associated with that error event;
- $N(w,i)$ expected number of error events beginning at a fixed symbol that cause i information bit errors and have Euclidean distance w.

The expectation is over the possible transmitted information sequences. The transfer function $T(W,I)$ can be computed by summing the path metrics in an appropriate state diagram of all paths with a nonzero exponent of I that begin and end in the zero error-state.

The technique of computing the transfer function to provide union bounds on BER for trellis codes has been widely investigated [19]–[22]. The transfer function can be found using various state transition diagrams. For a trellis code with ν_e memory elements, Biglieri [19] described a general algorithm using a $2^{2\nu_e}$-state transition diagram. Rouanne and Costello [20] and Zehavi and Wolf [21] demonstrated that a 2^{ν_e}-state transition diagram is sufficient for quasi-regular codes. Wesel [22] proposed to compute the transfer function of any trellis code using a $2^{\nu_e+\nu_q}$-state transition diagram where ν_q is an integer between zero and ν_e. Kucukyavuz and Fitz proposed an alternative state reduction technique in [23]. For simplicity and generality, we use Biglieri's product-state approach [19] to investigate the robustness of BICM and TCM schemes under a single channel fade (prediction error). However, the application of the state reduction techniques discussed above is straightforward.

Using the bound $Q(t) \leq \left(e^{-t^2/2}\right)/2$, the transfer function of a rate-k/n trellis code can provide the well-known union bound on the BER for an AWGN channel as

$$P_e \leq \frac{1}{2k} \frac{\partial T(W,I)}{\partial I}\bigg|_{I=1, W=e^{-1/4N_0}} \tag{3}$$

where $T(W,I)$ is the standard transfer function (2) and $N_0/2$ is the variance per dimension of the AWGN. Tighter bounds on $Q(\cdot)$ can be applied to (3) if the free distance of the code including fading effects is known [21].

Biglieri [19] used the $2^{2\nu_e}$-state transition diagram to compute $T(W,I)$ for larger constellations by employing the state pair (encoder state, incorrect encoder state) (say $(\sigma, \tilde{\sigma})$). Call the $N_g = 2^{\nu_e}$ product states where $\sigma = \tilde{\sigma}$ the "good product states" and the remaining $N_b = 2^{2\nu_e} - 2^{\nu_e}$ product states the "bad product states."

Define $\mathbf{A}$ to be the $N_b \times N_b$ matrix of labels of the form $W^w I^i$ for transitions from one bad product state to another. Here w is the squared Euclidean distance between the correct and incorrect constellation points produced in the product state transition, and i is the number of information bit errors associated with that transition. Define $\mathbf{b}$ to be the $N_b \times N_g$ "tall" matrix of labels of the form $W^w I^i$ for transitions from a good product state to a bad product state and $\mathbf{c}$ to be the $N_g \times N_b$ "wide" matrix of labels of the form $W^w I^i$ for transitions from a bad product state to a good product state, and $\mathbf{d}$ be the $N_g \times N_g$ matrix of labels of the form $W^w I^i$ for transitions from one good product state to another. Then the equation for the transfer function T is

$$\begin{aligned} T =& \frac{1}{2^{\nu_e}}\left(\mathbf{1}^T\mathbf{d}\mathbf{1} + \sum_{j=0}^{\infty}\mathbf{1}^T\mathbf{c}\mathbf{A}^j\mathbf{b}\mathbf{1}\right) \\ =& \frac{1}{2^{\nu_e}}\left(\mathbf{1^T}\mathbf{d}\mathbf{1} + \mathbf{1^T}\mathbf{c}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}\mathbf{1}\right) \end{aligned} \tag{4}$$

where $\mathbf{1}$ is the N_g-element column-vector of ones and $\mathbf{I}$ is the $N_b \times N_b$ identity matrix. In this case, the derivative $\partial T(W, I)/\partial I$ in (3) is

$$\begin{aligned} \frac{\partial T}{\partial I} =& \frac{1}{2^{\nu_e}}\Big(\mathbf{1^T}\mathbf{d}'\mathbf{1} + \mathbf{1^T}\mathbf{c}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}'\mathbf{1} \\ & + \mathbf{1^T}\mathbf{c}'(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}\mathbf{1} \\ & + \mathbf{1^T}\mathbf{c}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{A}'(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}\mathbf{1}\Big) \end{aligned} \tag{5}$$

where $\mathbf{d}'$, $\mathbf{b}'$, $\mathbf{c}'$, and $\mathbf{A}'$ indicate the element-by-element partial derivative with respect to I. Each term in the derivative of (5) counts information bit errors on only one trellis branch. This derivative effectively fixes attention on one trellis branch, counting the information bit errors occurring on that branch caused by all error events. For example, $\mathbf{1^T}\mathbf{c}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}'\mathbf{1}$ counts information bit errors for error events that begin with the trellis branch of interest.

To incorporate the effect of a single fade Y (i.e., the result of an incorrect prediction) in an AWGN channel for the trellis branch of interest at the transmitted lth symbol period time, replace $\mathbf{d}'$, $\mathbf{b}'$, $\mathbf{c}'$, and $\mathbf{A}'$ in (5) with $\mathbf{d}'_Y$, $\mathbf{b}'_Y$, $\mathbf{c}'_Y$ and $\mathbf{A}'_Y$, respectively, where in every element involving W in $\mathbf{d}'_Y$, $\mathbf{b}'_Y$, $\mathbf{c}'_Y$, and $\mathbf{A}'_Y$, the exponent of W is multiplied by the variable scale factor Y^2 instead of 1 ($|X(t)| = 1$ in an AWGN channel). For example, W^2 is replaced by W^{2Y^2}. Thus, for a trellis code under a single prediction error, the derivative in (3) becomes

$$\begin{aligned} \frac{\partial T}{\partial I} =& \frac{1}{2^{\nu_e}}\Big(\mathbf{1^T}\mathbf{d}'_\mathbf{Y}\mathbf{1} + \mathbf{1^T}\mathbf{c}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}'_\mathbf{Y}\mathbf{1} \\ & + \mathbf{1^T}\mathbf{c}'_\mathbf{Y}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}\mathbf{1} \\ & + \mathbf{1^T}\mathbf{c}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{A}'_\mathbf{Y}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}\mathbf{1}\Big). \end{aligned} \tag{6}$$

Therefore, by using (3) and (6), the transfer function bound on BER for TCM codes under a single prediction error can be obtained.

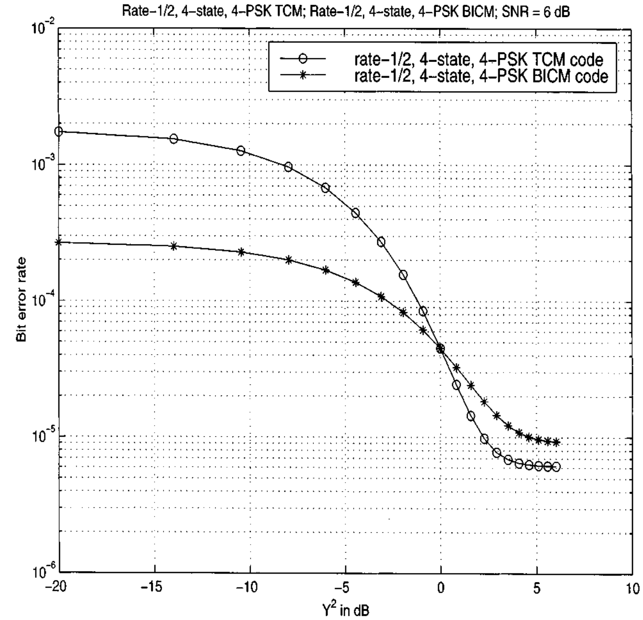

Fig. 2. Transfer function union bounds of the BER at the faded symbol for a rate-1/2, 4-state, 4-PSK BICM code and a rate-1/2, 4-state, 4-PSK TCM code under a single fade Y (or prediction error). The SNR for nonfaded symbols is 6 dB.

Similarly, for a rate-k/n BICM code, the transfer function T is [14]

$$\begin{aligned} T(W_1, \ldots, W_n, I) = & \sum_{w_1 \ldots w_n}\sum_i N(w_1, \ldots, w_n, i) I^i \\ & \times \prod_{j=1}^{n} W_j^{w_j} \end{aligned} \tag{7}$$

where $N(w_1, w_2, \ldots, w_n, i)$ is the number of error events with Hamming weight w_j associated with the jth most significant codeword bit before the interleaver, and i is the number of information bit errors associated with those error events.

Based on Caire *et al.*'s error-probability analysis for a single interleaver processing all bits [15], the union bound on BER for a rate-k/n BICM code over an AWGN channel is computed as

$$P_e \le \frac{1}{2k}\frac{\partial T(W_1, \ldots, W_n, I)}{\partial I}\bigg|_{I=1, W_j = N_{\min}(1)\exp\left(-d_{\min}^2/4N_0\right)} \tag{8}$$

where $N_{\min}(1)$ is a constant value and depends only on the constellation set and on the labeling [15], and $d_{\min}$ is the minimum Euclidean distance of the constellation set.

Similar to the discussion for the transfer function bound on BER for a trellis code under a single fade Y, the representation of the transfer function bound on BER for a rate-k/n BICM code under a single fade can be obtained from (6) and (8). Note that, for a BICM code, the incorrect prediction Y affects only one coded bit in each of n error events since every coded bit is interleaved by the ideal interleaver. Here the average case is considered; in other words, the fade Y is equally likely to affect each coded bit position in the trellis branch of interest in the associated error events. Specifically in $\mathbf{A}'$, $\mathbf{b}'$, $\mathbf{c}'$, and $\mathbf{d}'$, replace $\prod_{j=1}^{n} W_j^{w_j}$ with $1/n\sum_{p=1}^{n}\left(\prod_{j=1, j\neq p}^{n} W_j^{w_j}\right) W_p^{w_p Y^2}$

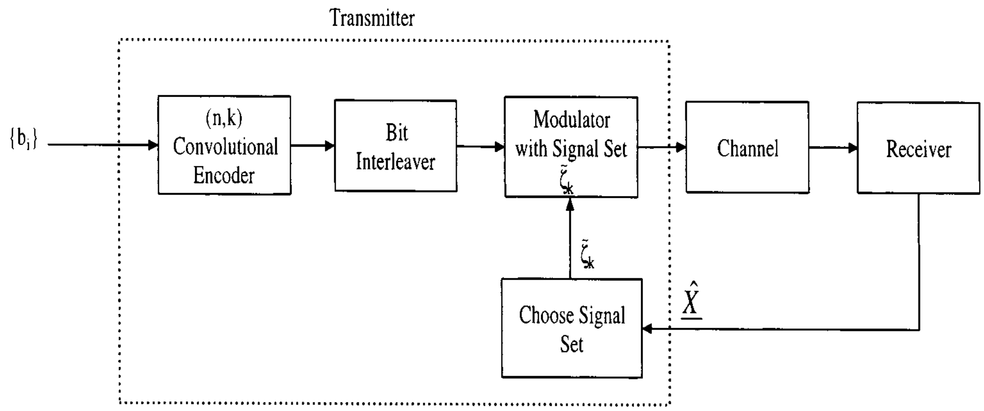

Fig. 3. The adaptive BICM paradigm.

to produce the $\mathbf{A}'_Y$, $\mathbf{b}'_Y$, $\mathbf{c}'_Y$, and $\mathbf{d}'_Y$ for application of (6) to the BICM case.

Fig. 2 shows the result of applying these two modified transfer function bounds to TCM and BICM, respectively, each with a single fade Y at the lth transmitted symbol period. The figure compares a rate-1/2, 4-state, 4-PSK TCM with a rate-1/2, 4-state, 4-PSK BICM. These two codes were selected because they provide the same performance on the AWGN channel. Both codes have a BER of 4×10^{-5} at 6 dB, which is the SNR examined in Fig. 2.

To interpret Fig. 2 in the context of adaptive coding, consider the estimate for the lth symbol to have been normalized to 1. Y^2 (the x axis) is the square of the actual fading value (given in decibels). When the decibel value of Y^2 is zero, the estimate is perfectly correct. When the decibel value of Y^2 is negative, the estimate is optimistic. This is the situation where adaptive-TCM is known to fail, and Fig. 2 illustrates how BICM provides significantly better performance (more robustness) in this region. The reason for the better performance of BICM in this region is that TCM sees the full effect of the small value of Y^2 on a single trellis branch while BICM spreads the effect of this prediction error over n well-separated trellis branches.

When the decibel value of Y^2 is positive, the estimate is pessimistic. In this region, TCM has a slight advantage over BICM. The explanation of the TCM advantage in this region is actually the same as the reason for its poor performance for the optimistic prediction error. In this situation, the pessimistic prediction causes one constellation to have a larger-than-expected minimum distance. With TCM, the full effect of this larger distance is seen on a single trellis branch. For BICM, this effect is spread over n well-separated branches, causing a smaller decrease in BER at each branch. Hence, BICM weakens the effect of prediction errors whether that effect is detrimental or beneficial to the BER.

IV. Adaptive BICM Paradigm

In the past, adaptive signaling generally has been performed using TCM [3]–[5], [10]. For these schemes, the number of information bits entering the encoder is held fixed for all signal sets and adaptivity is provided by changing the number of uncoded bits according to the estimate of the current channel fading value. It is shown in [4], [5] that TCM schemes which do not take into account the variation of the wireless channel over time work acceptably only when the product of the delay between channel estimation and data transmission and the Doppler frequency is very small. Thus, in general, adaptive TCM schemes should explicitly take into account the variation of the fading channel over time. In this paper, the design of adaptive BICM will be analyzed using both types of constructions: the first assumes that the current channel fading values are known perfectly at the transmitter (this construction will be called the "deterministic design" from hereon); the second construction explicitly takes into account the variation of the channel over time. The adaptive BICM paradigm that will be followed is shown in Fig. 3. The output coded sequence from the convolutional encoder is first interleaved by the bit interleaver. The signal constellation $\tilde{\zeta}_k$ for the kth signaling interval is chosen from the allowable signal sets $\zeta_1, \zeta_2, \ldots, \zeta_L$, where the signal constellation ζ_{l+1} has a size larger than that of ζ_l, based on the information about the current fading provided by $\underline{\hat{X}}$. Let M denote the size of the chosen constellation $\tilde{\zeta}_k$. Then, $\log_2 M$ bits are taken from the interleaver and used to choose a signal from $\tilde{\zeta}_k$. It is conjectured in [15] that Gray labeling maximizes the BICM capacity over Rayleigh fading channels; henceforth, in the design process Gray labeling will be used to map the output of the bit interleaver onto signals. Finally, the resulting signal is transmitted over the channel. Note that there are no uncoded information bits sent through the channel. The metric generation of this adaptive scheme for a given signal set is the same as given in [15].

There are two design issues that must be considered: how to choose the convolutional encoder, and how to choose the signal set $\tilde{\zeta}_k$ based on $\underline{\hat{X}}$. The first issue is easily addressed; for a fixed code rate and constraint length, a convolutional code of maximal free Hamming distance is employed. For the second issue, two types of steps can be taken. For the deterministic design, it is assumed that the current channel fading value is predicted perfectly from the outdated fading estimates. For the second design, the design rules given in [4] and [5] for adaptive TCM will be modified. The next section considers the design issues for both types of constructions.

A. Design Rules

For the paradigm, a nominal nonadaptive BICM scheme as in [15] is assumed that employs signal set ζ_0 and provides the desired bit error probability P_b at some signal-to-noise ratio $(\mathrm{SNR})_0$ on a Rayleigh fading channel. The same trellis structure of the nominal scheme is then used for the adaptive design. Let u_l be the minimum normalized Euclidean distance squared between any two points in the signal constellation ζ_l. In other words, letting $d_E(x,v)$ be the Euclidean distance between signals x and v yields

$$u_l = \min_{x,v,x\neq v} \frac{d_E^2(x,v)}{E_l} \tag{9}$$

where E_l is the average energy of the signal set ζ_l.

1) Deterministic Design: When the current channel fading value $Y = y$ is assumed to be known exactly from $\underline{\hat{X}}$, the following rule can be employed: for the kth symbol, transmit a signal from the signal set $\tilde{\zeta}_k$, where it is the largest signal set ζ_l such that $y^2 u_l \geq u_0$, where u_0 is the minimum normalized Euclidean distance squared between any two points in the nominal signal constellation. Note that this technique allows the system to refrain from signaling during deep fades, thus alleviating a key inhibiting factor to fading channel transmission.

2) Stochastic Design: The design of adaptive TCM for time-varying channels [4], [5] consists of two key considerations: realizing that the conditional probability distribution of the current fading value varies from Rayleigh to strongly Rician, and using this observation to conclude that the minimum intersubset and intrasubset normalized Euclidean differences, which are defined as the ability to decide between symbols from different subsets and within the same subset, respectively, should be maintained separately. This design recipe will be adapted to BICM in the following way: since there are no uncoded bits in the adaptive BICM scheme, there are no intrasubset differences to be maintained. The only constraint is on the intersubset distance, which for BICM can be redefined as the minimum normalized Euclidean distance squared between any two signals in a given signal constellation, namely u_l for any constellation ζ_l. With this definition, the design reduces to that of [4] and [5], and the formula given for the intersubset difference can be used to select the signal constellations according to the outdated channel estimate. That is, letting $h = |X(kT_s - \tau_1)|$, u_l must guarantee

$$\sup_{\rho \geq \rho_{\min}} \int_0^\infty \exp\left(-y^2 u_l \frac{E_s}{8N_0}\right) p_{Y|h}(y|h)\,dy \leq D_0(0, u_0 E_s) \tag{10}$$

in order to approximately maintain the bit error probability of the nominal scheme. E_s is the average transmitted energy. Here, $D_0(x,v)$ is the Bhattacharyya bound [24] on the error probability of choosing x when v is correct for the nominal code. For example, $D_0(x,v) = 1/\left(1 + (d_E^2(x,v)/4E_s)(\mathrm{SNR})_0\right)$ if the nominal code is designed for a Rayleigh channel. The evaluation of the left-hand side of (10) in terms of h, the average transmitted SNR E_s/N_0, and $\rho_{\min}$ is given in [4] and [5]. Making the intersubset difference guarantee (10) with equality for each l yields h_l, the threshold such that for $h \geq h_l$, ζ_l can be employed.

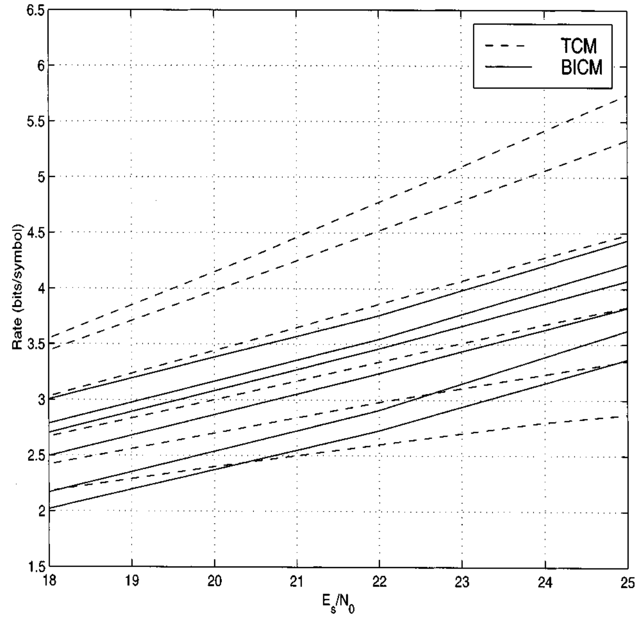

Fig. 4. The solid lines show the simulated rate of adaptive BICM using the stochastic design, for the $\rho_{\min}$ values 1.0, 0.97, 0.95, 0.90, 0.80, and 0.70 from top to bottom, respectively. The simulated P_b for each data point is less than the target 10^{-5}, except for $\rho_{\min} = 0.70$ at 18 dB, for which P_b is 1.1×10^{-5}. The corresponding ($r = 2$ bits/symbol) nonadaptive BICM achieves $P_b = 10^{-5}$ at $E_s/N_0 = 18$ dB. The dashed lines show the simulated rate of adaptive TCM of [4], [5] for the $\rho_{\min}$ values 1.0, 0.99, 0.97, 0.95, 0.93, and 0.90 from top to bottom, respectively. The corresponding ($r = 2$ bits/symbol) nonadaptive I-Q TCM [26] achieves $P_b = 10^{-5}$ at $E_s/N_0 = 17.5$ dB.

The rate of the above design schemes is limited due to the fact that for all h such that $h_l < h < h_{l+1}$, the estimate is more favorable than that required to use ζ_l but not favorable enough to use ζ_{l+1}. This problem can be overcome by employing energy adaptation as in [4], [5]. First, a signal set is chosen according to the above given design method with no energy adaptation. Then (10) is used to decide the minimum energy required to maintain D_0 given the channel estimate h. This minimum energy is then employed for the current symbol, and the remaining energy is saved for the next symbol. This algorithm will be used for all of the results below.

V. Numerical Results and Analysis

For the nominal BICM scheme, the two bits per 8-PSK symbol 8-state code from [15], which operates at the desired bit error probability $P_b = 10^{-5}$ at average transmitted SNR $(\mathrm{SNR})_0 = 18$ dB per PSK symbol on a Rayleigh fading channel, is employed. This implies that a rate-2/3 8-state convolutional encoder of maximum free distance is employed in the adaptive scheme. First, consider the stochastic design. The Bhattacharyya bound D_0 on the right hand side of (10) is found using the nominal 8-PSK constellation of [15]. Let the candidate signal sets $\zeta_1, \zeta_2, \ldots, \zeta_8$ be the set of M-ary QAM constellations, $M = 2, 4, 8, 16, 32, 64, 128, 256$ ($M = 2$ implying BPSK), respectively. Note that for $M = 8, 32,$ and 128, Gray-labeling is not possible; for these constellations, quasi-Gray labeling [25], which minimizes the number of signals for which the Gray-labeling condition is not satisfied, is employed. Numerical results, which employ an energy margin

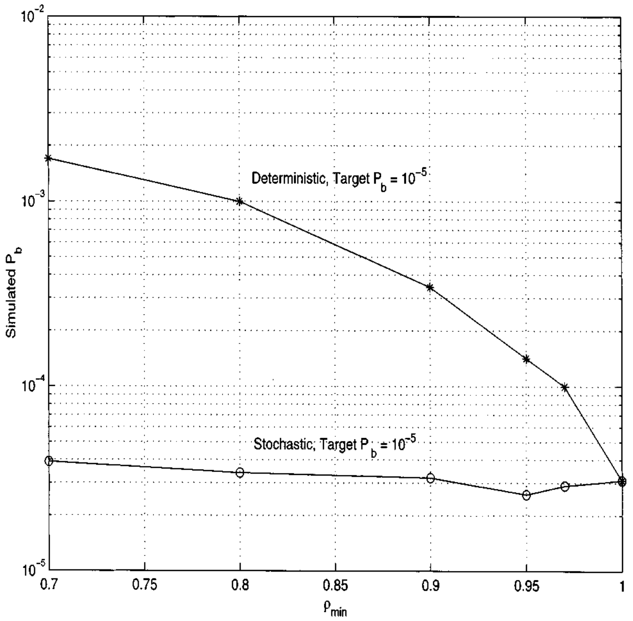

Fig. 5. Simulated probability of bit error of adaptive BICM designs versus $\rho_{\min}$ at average transmitted SNR $E_s/N_0 = 18$ dB per QAM symbol. Both designs use rate 2/3, 8-state convolutional encoders of maximum free distance. No energy margin is employed.

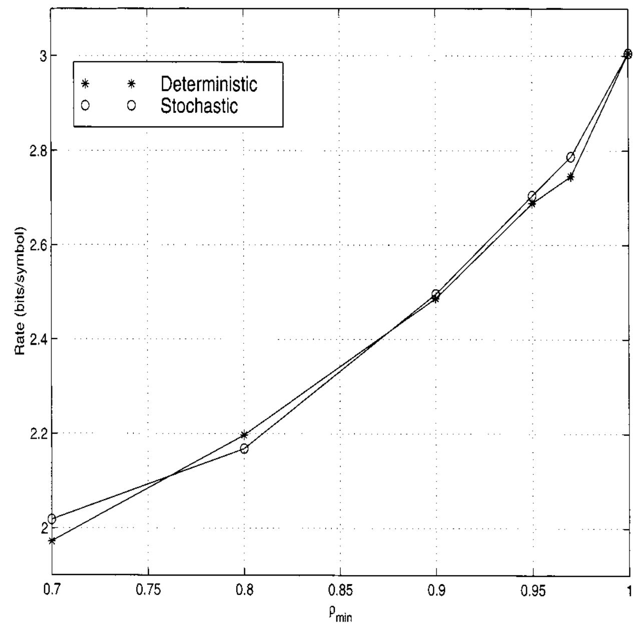

Fig. 6. Simulated rate of adaptive BICM schemes versus $\rho_{\min}$ at average transmitted SNR $E_s/N_0 = 18$ dB per QAM symbol. Both designs use rate 2/3, 8-state convolutional encoders of maximum free distance. Energy margin is employed in both schemes to achieve target $P_b = 10^{-5}$.

as detailed below, for the rate of the adaptive BICM scheme are shown in Fig. 4. Fig. 4 also displays the simulated rates achieved by the adaptive TCM of [4], [5]. At high values of $\rho_{\min}$, the TCM scheme of [4], [5] is superior to the proposed BICM scheme as expected. For small values of $\rho_{\min}$ (implying an increase in system user mobility), adaptive BICM outperforms the nonadaptive schemes, while adaptive TCM schemes are below the rate achieved by their nonadaptive counterparts.

Fig. 5 displays the probability of bit error of the deterministic design and of the stochastic design as a function of $\rho_{\min}$ at 18-dB average transmitted SNR. As can be seen, due to prediction errors, the deterministic design misses the desired bit error rate by a significant amount. The stochastic design misses the target P_b as well, though by a much smaller amount. The reason that it misses the target is that the Bhattacharyya bound based only on the distance between the two nearest signals in the constellation, which is used to find the thresholds per (10), is not accurate at these SNRs. Note that each of these two systems is operating at the rate produced by its constellation size selection thresholds h_l. The deterministic design has more aggressive thresholds; it achieves a higher rate but has a poor probability of bit error performance.

An energy margin can be employed in both schemes to reach the desired BER; that is, in (10), instead of the actual average transmitted SNR, a smaller SNR value is used to find the thresholds. Doing so increases the thresholds, thus decreasing the rate and the probability of bit error of the system. Employing appropriate energy margins for both schemes to reach the desired probability of bit error $P_b = 10^{-5}$ at 18-dB average transmitted SNR results in the rates displayed in Fig. 6. As can be seen, there is no extra bandwidth efficiency gained by considering the prediction errors explicitly, which is in contrast to what was observed in [4], [5] for adaptive TCM schemes. This can be explained as follows. When there are parallel branches in the trellis, it is shown in [4], [5] that the intrasubset minimum normalized Euclidean distance squared of the possible constellations should satisfy an equation similar to (10) except that the right hand side is changed to P_b, the desired probability of bit error for the system. The threshold on h above which a signal set can be employed then becomes the maximum of the thresholds found by the two equations using intersubset and intrasubset distances. The Bhattacharyya parameter on the right-hand side of (10) is usually quite high (in our case, on the order of 10^{-1}) as compared to the BER requirement for the uncoded bits (in our case, on the order of 10^{-5}). Recall that the deterministic design assumes that $\rho_{\min} = 1$, while actually the fading channel has a $\rho_{\min}$ value smaller than 1. As an example, consider the above design with the actual $\rho_{\min}$ equal to 0.90. For this BICM scheme, the right-hand side of (10) is $D_0 = 0.0976$ at 18-dB average transmitted SNR. In Fig. 7, the left-hand side of (10) between the range 0.097 and 0.099 at 18-dB average transmitted SNR is plotted versus h for all possible constellations (M-QAM, $M = 2^l$, $l = 1, 2, \ldots, 8$) used in the adaptive system for $\rho_{\min} = 0.90$. The intersection of a curve with the $D_0 = 0.0976$ line determines the threshold above which the corresponding signal set can be employed. The deterministic design uses the plots for $\rho_{\min} = 1$ to determine the thresholds, when it should use the ones for which $\rho_{\min} = 0.90$. The addition of an energy margin results in an almost exact alignment with the actual curves; that is, the threshold boundaries of the signal sets for the deterministic case and for the actual case are nearly at the correct ratio. Hence, by adding an energy margin, the performance of both schemes can be made virtually the same. However, if the same curves are observed in the range of 10^{-5} (corresponding to a $P_b = 10^{-5}$ requirement), it can be seen that regardless of the energy margin employed, the curves for the determin-

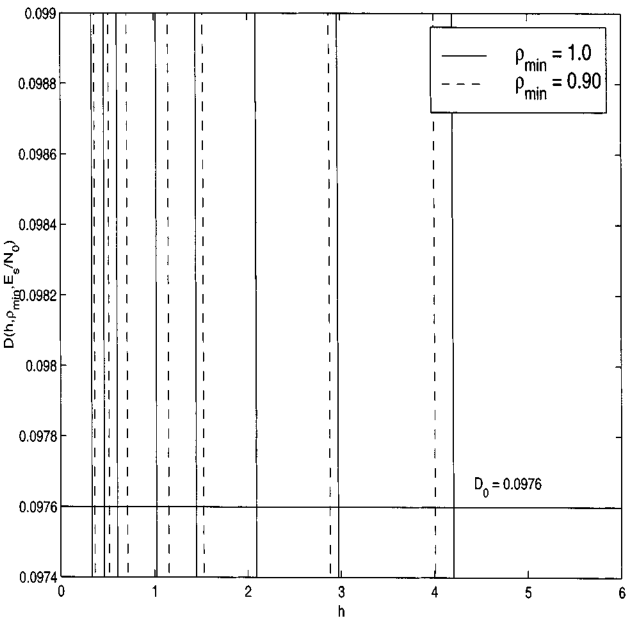

Fig. 7. Left hand side of (10) versus h for all possible constellations at average transmitted SNR $E_s/N_0 = 18$ dB per QAM symbol. The solid curves are for $\rho_{\min} = 1$ starting at left from BPSK, going to 256-QAM. An energy margin of 1.5 dB is employed for $\rho_{\min} = 1$ case. Dashed curves correspond to the $\rho_{\min} = 0.90$ case.

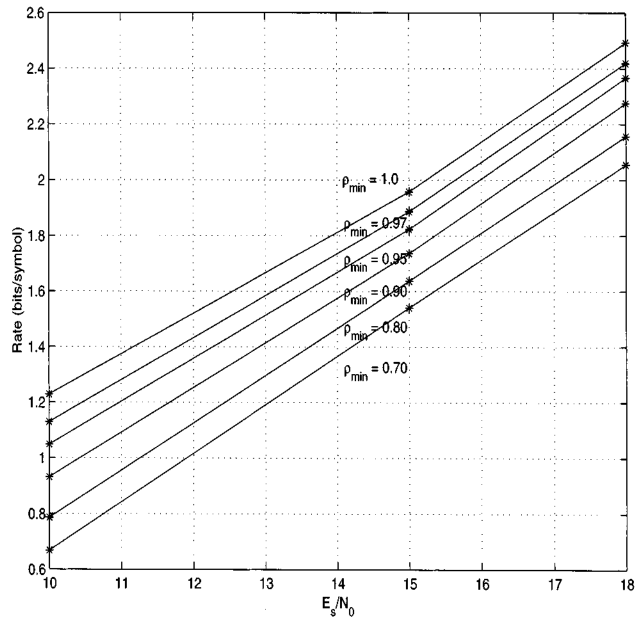

Fig. 8. Simulated rate versus average transmitted SNR of adaptive BICM using rate 1/2 8-state convolutional encoder using the stochastic design, for several $\rho_{\min}$ values. The simulated P_b for each data point is less than the target 10^{-5} for all points. The corresponding ($r = 1$ bits/symbol) nonadaptive BICM achieves $P_b = 10^{-5}$ at $E_s/N_0 = 10$ dB.

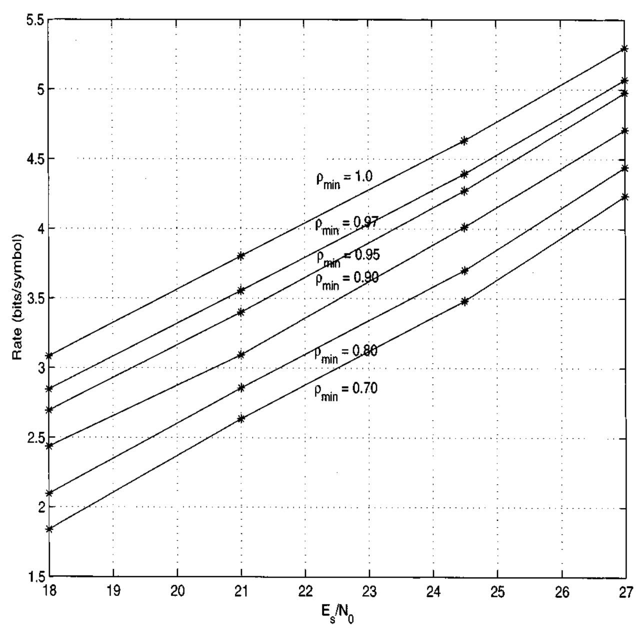

Fig. 9. Simulated rate versus average transmitted SNR of adaptive BICM using rate 3/4 8-state convolutional encoder using the stochastic design, for several $\rho_{\min}$ values. The simulated P_b for each data point is less than the target 10^{-5} for all points. The corresponding ($r = 3$ bits/symbol) nonadaptive BICM achieves $P_b = 10^{-5}$ at $E_s/N_0 = 21$ dB.

istic case and the actual case can never be aligned. Therefore, in this case significant prediction error occurs whenever thresholds derived with the deterministic design are used instead of the actual ones; thus, adaptive TCM schemes, which employ uncoded bits that must meet the $P_b = 10^{-5}$ requirement, must take into account the variation of the channel over time. However, for BICM, which uses only the intersubset distances, an energy margin can always be employed to align the deterministic thresholds with the correct ones, thus removing the need to consider the variation of the channel over time explicitly.

Two other nominal schemes with different encoder rates have also been considered for the adaptive BICM system. The first nominal scheme is the 1 bit per 4-QAM code, which achieves the target probability of bit error $P_b = 10^{-5}$ at 10-dB average transmitted SNR. A rate-1/2 8-state convolutional encoder of maximum free distance is used for this scheme. The second nominal scheme uses a rate-3/4 8-state convolutional encoder with a 16-QAM constellation, and achieves target $P_b = 10^{-5}$ at 21-dB average transmitted SNR. Figs. 8 and 9 show the achieved rates by designing adaptive BICM based on these schemes as a function of average transmitted SNR for several $\rho_{\min}$ values. From these figures, together with Fig. 4, several conclusions can be made. First, compare the rates of simulated adaptive BICM schemes at 18-dB average transmitted SNR for different $\rho_{\min}$ values. For $\rho_{\min} \geq 0.70$, although all three schemes are almost always better than the nonadaptive design, which achieves a rate of 2 bits/symbol at 18-dB average transmitted SNR, the best code to use at 18 dB (or any other given average transmitted SNR) changes according to the value of $\rho_{\min}$. Thus, it can be concluded that there is no single nominal scheme that achieves the highest rate for all channel $\rho_{\min}$ values at a given average transmitted SNR. Fig. 10 plots the rates of these schemes as a function of $\rho_{\min}$ at 18-dB average transmitted SNR. The plot reveals that at a given SNR, using a smaller rate convolutional encoder is better for small $\rho_{\min}$ values. However, when the value of $\rho_{\min}$ is in the larger part of its range, a higher code rate gives higher bandwidth efficiency. It can also be seen from the figure that the slopes of the curves increase with increased encoder rate. Both of these results are expected: when $\rho_{\min}$ is relatively small, implying that the conditional channel is nearly Rayleigh, the higher time diversity of the 8-state codes of lower rate in

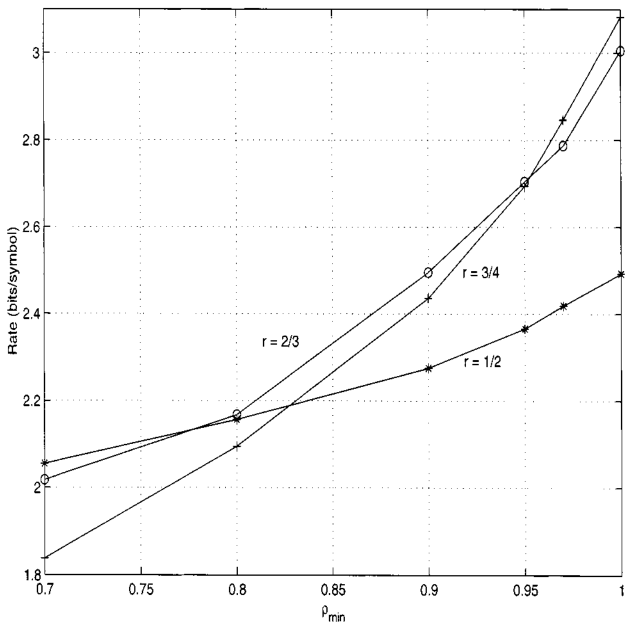

Fig. 10. Simulated rates of adaptive BICM schemes using the stochastic design, as a function of $\rho_{\min}$ at 18-dB average transmitted SNR.

Fig. 10 allows the correction of more prediction errors and thus the achievement of a higher bandwidth efficiency. However, for a high $\rho_{\min}$ value, prediction errors are less prevalent and the channel is utilized most efficiently by using a higher rate code.

VI. Conclusion

Although adaptive TCM systems demonstrate large bandwidth efficiency gains over their nonadaptive counterparts in wireless systems that support users with low mobility, their rates plummet considerably as the mobility of the users increases. Thus, to extend the applicability of adaptive systems, a new adaptive coding paradigm based on bit-interleaved coded modulation has been introduced for bandwidth-efficient modulation. It is shown that the structure of BICM is more robust to prediction errors than TCM, thus motivating the use of adaptive BICM systems in systems where outdated channel estimates can be inaccurate. Since adaptive BICM is robust to prediction errors, the explicit consideration of variation of the channel over time is not necessary; assuming the current channel estimate is known exactly from the outdated estimates and then employing energy margin leads to a negligible decrease in performance. The numerical results show that adaptive BICM schemes extend the applicability of adaptive signaling to systems with more highly mobile users than previously considered.

References

[1] A. Goldsmith, "Capacity and dynamic resource allocation in broadcast fading channels," in *Proc. Allerton Conf. Commun., Contr. Comput.*, 1995, pp. 915–924.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. ICC '93*, May 1993, pp. 1064–1070.

[3] A. Goldsmith, "Variable-rate coded M-QAM for fading channels," in *Proc. IEEE Global Commun. Conf.—Commun. Theory Miniconf.*, 1994, pp. 186–190.

[4] D. Goeckel, "Robust adaptive coding for time-varying channels with delayed feedback," in *Proc. Allerton Conf. Commun., Contr. Comput.*, 1997.

[5] ——, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, pp. 844–855, June 1999.

[6] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[7] D. Goeckel, "Adaptive coding for fading channels using outdated fading estimates," in *Proc. 1998 IEEE 48th Vehicular Tech. Conf*, May 1998, pp. 1925–1929.

[8] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.

[9] B. Vucetic, "An adaptive coding scheme for time varying channels," *IEEE Trans. Commun.*, vol. 39, pp. 653–663, May 1991.

[10] S. Alamouti and S. Kallel, "Adaptive trellis coded multiple-phase-shift keying for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 2305–2314, June 1994.

[11] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[12] M. Alouini and A. Goldsmith, "Adaptive M-QAM modulation over Nakagami fading channels," in *Proc. IEEE Global Communications Conf.—Communication Theory Miniconf.*, Nov. 1997, pp. 218–223.

[13] J. Cavers, "Variable-rate transmission for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 20, pp. 15–22, Feb. 1972.

[14] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.

[15] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.

[16] C. Schlegel and D. Costello Jr, "Bandwidth efficient coding for fading channels: Code construction and performance analysis," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1356–1368, Dec. 1989.

[17] S. Wilson and Y. Leung, "Trellis-coded phase modulation on Rayleigh channels," in *Proc. Int. Conf. Commun.*, 1987, pp. 739–743.

[18] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. COM-11, pp. 360–393, Dec. 1963.

[19] E. Biglieri, "High-level modulation and coding for nonlinear satellite channels," *IEEE Trans. Commun.*, vol. 32, pp. 616–626, May 1984.

[20] M. Rouanne and D. Costello Jr, "An algorithm for computing the distance spectrum of trellis codes," *J. Select. Areas Commun.*, vol. 7, pp. 929–940, Aug. 1989.

[21] E. Zehavi and J. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. Inform. Theory*, vol. 33, pp. 196–202, Mar. 1987.

[22] R. Wesel, "Reduced complexity trellis code transfer function computation," in *Commun. Theory Miniconf. (in conjunction with ICC99)*, June 6–10, 1999.

[23] D. Kucukyavuz and M. P. Fitz, "New views of transfer function based performance analysis of coded modulations," in *33th Asilomar Conf. Signals, Syst. Comput.*, Oct. 1999.

[24] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[25] X. Liu and R. Wesel, "Profile optimal 8-QAM and 32-QAM constellations," in *36th Annu. Allerton Conf. Commun., Contr., Computing*, Sept. 1998.

[26] S. Al-Semari and T. Fuja, "I-Q TCM: Reliable communication over the Rayleigh fading channel close to the cutoff rate," *IEEE Trans. Inform. Theory*, vol. 43, pp. 250–262, Jan. 1997.

Pınar Örmeci received the B.S. degree in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 1998, and the M.S. degree from the University of Massachusetts, Amherst, in 2000.

From 1998 to 2000, she was a Research Assistant with the Electrical and Computer Engineering Department, University of Massachusetts, Amherst. Her graduate research focused on digital communication systems for wireless applications. Since September 2000, she has been working in the Advanced Technology Group of Narad Networks, Westford, MA. She is currently involved in the development of novel modem architectures.

Xueting Liu received the B.S. degree in automatic control engineering from Central South University of Technology, Changsha, China, in 1992, and the M.S. degree from the Institute of Automation, Chinese Academy of Sciences,Beijing, China, in 1995. She is currently working toward the Ph.D. degree with the Electrical Engineering Department, University of California at Los Angeles.

Her research interests include the design of families of rate-compatibile, unequal-error protection trellis codes and their application in speech transmission, the investigation of constellation labeling paradigm, and adaptive trellis code design for fading channels

Dennis L. Goeckel received the B.S.E.E. degree from Purdue University, West Lafayette, IN, in 1992, and the M.S.E.E. and Ph.D. degrees, both in electrical engineering, from the University of Michigan, Ann Arbor, in 1993 and 1996, respectively.

From 1987 to 1992, he worked at Sundstrand Corporation. From 1992 to 1996, he was a National Science Foundation Graduate Fellow at the University of Michigan. In September 1996, he assumed his current position as an Assistant Professor with the Electrical and Computer Engineering Department, University of Massachusetts, Amherst. His current research interests are in the design of digital communication systems, particularly for wireless communications applications.

Dr. Goeckel is the recipient of a 1999 CAREER Award from the National Science Foundation. Currently, he is an Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: WIRELESS COMMUNICATIONS SERIES.

Richard D. Wesel (S'91–M'96) received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1989 and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1996.

From 1989 to 1991, he was with AT&T Bell Laboratories, where he worked on nonintrusive measurement and adaptive correction of analog impairments in AT&T's long-distance network and the compression of facsimile transmissions in these areas. Since 1996, he has been an Assistant Professor with the Electrical Engineering Department, University of California at Los Angeles. His research interests include coded modulation, communication theory, and data fusion in distributed sensor networks.

Dr. Wesel received a National Science Foundation CAREER Award to pursue research on robust and rate-compatible coded modulation. He also received an Okawa Foundation Award in 1999 for his research in information and telecommunications. In 1999, h e was appointed to the editorial staff of the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of coding and coded modulation.