

# Joint Iterative Channel Estimation and Decoding in Flat Correlated Rayleigh Fading

Christos Komninakis, *Student Member, IEEE*, and Richard D. Wesel, *Senior Member, IEEE*

**Abstract**—This paper addresses the design and performance evaluation with respect to capacity of M-PSK turbo-coded systems operating in frequency-flat time-selective Rayleigh fading. The receiver jointly performs channel estimation and turbo decoding, allowing the two processes to benefit from each other. To this end, we introduce a suitable Markov model with a finite number of states, designed to approximate both the values and the statistical properties of the correlated flat fading channel phase, which poses a more severe challenge to PSK transmission than amplitude fading. Then, the Forward–Backward algorithm determines both the maximum *a posteriori* probability (MAP) value for each symbol in the data sequence and the MAP channel phase in each iteration. Simulations show good performance in standard correlated Rayleigh fading channels. A sequence of progressively tighter upper bounds to the capacity of a simplified Markov-phase channel is derived, and performance of a turbo code with joint iterative channel estimation and decoding is demonstrated to approach these capacity bounds.

**Index Terms**—Capacity, flat Rayleigh fading, Markov channels, turbo codes.

## I. INTRODUCTION

CODING can help combat the adversities of multipath propagation and scattering in a wireless fading channel. In particular, when the transmission rate is relatively slow, the channel typically exhibits little frequency selectivity but often significant time variation within a packet, depending on the relative velocity between the transmitter and receiver. This paper concentrates on the design of efficient joint turbo-coding [1] and channel estimation schemes that achieve reliable communication in such frequency-flat time-selective Rayleigh fading channels, arising for instance in the reverse link of CDMA systems.

Early coding work for flat Rayleigh fading, comprehensively summarized in [2], assumed the availability of perfect channel state information (CSI) at the receiver and designed trellis codes accordingly. A more sophisticated coding strategy named Bit-Interleaved Coded Modulation (BICM—in [3]) offers better performance in flat fading, but also relies on CSI. In [4] and later [5] the emphasis is on joint estimation and decoding for Markov channels via decision-feedback-aided recursive

channel estimators. Although proven to achieve the capacity of the Markov channel in the absence of error propagation, those receivers are vulnerable to error propagation and unreliable when the channel quality degrades. Recently, the advent of turbo codes [1], whose impressive error-correcting capabilities in AWGN permit operation at low signal-to-noise ratio (SNR), coupled with new applications in channels with higher Doppler calls for the problem to be considered anew. The reason is that traditional channel estimation techniques, such as a decision-directed phase-locked loop (PLL) or pilot symbol assisted modulation (PSAM, in [6]) may fail to cope with a high-noise, fast changing channel.

In this context, research has explored ways to utilize the enormous potential of the Forward–Backward algorithm [7]—the basic tool for turbo-decoding—in concatenating two or more estimation schemes [8], and thus combating distortion due to channels more severe than AWGN. In [9] turbo-code operation in a rough approximation of a fading channel with a two-state Markov model (with a “Good” and a “Bad” state) was presented. The ISI channel was treated in [10] by exchanging soft information between a soft “equalizer” and “decoder,” while [11] deals with the frequency-selective fading channel of GSM or IS-54 applications. Also, [12] and [13] explored pilot-symbol-assisted turbo codes in flat fading, where the pilot-averaging filters produce refined channel estimates as the turbo-decoder iterations proceed. Part of the present paper may be considered an extension of [9] to more complex Markov channel models, representing the real world fading more accurately. Also, it may be considered similar to [12] and [13] with the Forward–Backward algorithm used instead of optimum filtering to perform channel estimation.

To facilitate the use of the Forward–Backward algorithm, a finite-state Markov model approximates the values as well as the statistical properties of the channel phase, similarly to [14], where a Markov model was proposed for the amplitude fading of a phase-coherent Rayleigh fading channel. Our Markov channel model is designed to match the phase process, a choice justified from simulation results which show that amplitude fading with PSK turbo codes is adequately estimated from the received amplitude with a simple affine estimator. The approximation of Clarke’s fading [15], as described in Section II, with a discrete Markov process has been shown to capture most (but not all) of the dynamics of the fading process in [16] (for amplitude fading there, but the result also carries over to the channel phase process).

The receiver algorithms proposed in this paper use the finite Markov model for the channel phase to integrate channel estimation and data decoding in an iterative fashion, possibly aided

Manuscript received May 1, 2000; revised November 5, 2000. This work was supported by Texas Instruments Incorporated, National Science Foundation Career Award CCR-9733089, and the Xetron Corporation. This paper was presented in part at Globecom 1999, Rio de Janeiro, Brazil, December 5–9, 1999, and in part at the 33rd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October 24–27, 1999.

The authors are with the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA 90095-1594 USA (e-mail: chkomn@ee.ucla.edu; wesel@ee.ucla.edu).

Publisher Item Identifier S 0733-8716(01)02554-9.

by pilot symbols periodically injected in the coded data stream. At each iteration, the soft-input/soft-output (SISO) modules [8] running the Forward–Backward algorithm produce and exchange soft information about the channel phase as well as the data, allowing the decoding and channel estimation procedures to benefit from each other. With this approach, as the reliability of the correct data increases with successive iterations of the decoding algorithm, essentially every coded symbol gradually comes to serve as a pilot symbol. Performance is checked both in flat Rayleigh fading and in a simplified finite-state Markov channel, for which we develop a sequence of progressively tighter bounds to capacity and show simulations that approach those bounds with reasonable complexity.

The organization of the paper is as follows. Section II discusses the channel model for flat correlated Rayleigh fading, followed by the Markov model for the phase process. Section III presents the algorithms for joint iterative decoding and channel estimation and their performance for various Doppler rates. Section IV derives capacity bounds for a simplified discrete Markov channel closely connected to the real-world fading channel of Section II and assesses the performance of the proposed iterative schemes in the light of those bounds. Finally, Section V concludes the paper.

## II. CHANNEL MODEL

A widely accepted statistical model for nondispersive wireless fading channels has been introduced in [17] and is commonly referred to as Clarke’s fading model. According to this model, after matched filtering and proper sampling, the discrete representation of the received signal at time  $t$  is

$$y_t = a_t \cdot x_t + n_t, \quad t = 0, 1, 2, \dots \quad (1)$$

where

- $x_t$  transmitted constellation point (in this paper from an  $M$ -PSK constellation);
- $\{n_t\}$  i.i.d. (white) complex Gaussian noise process, with variance  $\sigma^2 = N_o/2$  per dimension;
- $\{a_t\}$  correlated channel fading process, modeled as a circular complex Gaussian random process.

Assuming absence of line of sight and a continuum of scatterers in the vicinity of the omnidirectional mobile receiver antenna, we write  $a_t = X_t + jY_t = |a_t|e^{j\phi_t^a}$ , where  $\{X_t\}$  and  $\{Y_t\}$  are mutually uncorrelated zero-mean Gaussian processes, each with correlation properties determined by the Doppler frequency  $f_D$ ; see [18]

$$R_c(\tau) = E[X_t X_{t+\tau}] = E[Y_t Y_{t+\tau}] = \gamma^2 \mathcal{J}_0(2\pi f_D \tau) \quad (2)$$

where  $\mathcal{J}_0(\cdot)$  is the zero-order modified Bessel function of the first kind, and  $\gamma^2 = 0.5$  for normalized power. This autocorrelation gives rise to the well-known U-shaped normalized power spectral density in Jakes [15]

$$S_{xx}(f) = S_{yy}(f) = \frac{1}{\pi f_D T \sqrt{1 - \left(\frac{f}{f_D T}\right)^2}}. \quad (3)$$

If  $\{X_t\}$  and  $\{Y_t\}$  are zero-mean (worst case), the marginal distributions of  $|a_t|$  and  $\phi_t^a$  are Rayleigh and uniform respectively, hence the term “Rayleigh fading.”

Clarke’s model for the channel fading process is realistic and has been found to quite accurately match field measurements of physical channels. However, the non-Markovian autocorrelation properties of the amplitude process  $\{|a_t|\}$  and, mainly, the phase process  $\{\phi_t^a\}$ , which poses the primary problem in PSK transmission, are difficult to analyze and exploit with the Forward–Backward algorithm. Therefore, we derive a suitable finite-state Markov model for the channel phase, depending on the Doppler rate  $f_D T$ , similar to the one in [14] for the fading amplitude.

The receiver forms a  $K$ -state Markov model for the quantized version  $Q_t$  of the phase fading process  $\phi_t^a$ , where  $\{Q_t\}$ ,  $t = 0, 1, 2, \dots$  is a time-homogeneous, discrete-time, stationary Markov chain, taking values in the finite state space  $\mathcal{Q} = \{q_0, q_1, \dots, q_{K-1}\}$ , a set of “quantized channel phase distortion states”  $q_i$

$$q_i = \frac{2\pi i}{K}, \quad i = 0, 1, 2, \dots, K-1 \quad (4)$$

in the following fashion, where we introduce a quantization operator  $\Pi(\cdot)$ :

$$Q_t = q_i \Leftrightarrow \Pi(\phi_t^a) = q_i \Leftrightarrow \phi_t^a \in \left[ q_i - \frac{\pi}{K}, q_i + \frac{\pi}{K} \right).$$

The transition probabilities  $P_{q_i, q_j}$ ,  $i, j = 0, 1, \dots, K-1$  of the Markov chain are independent of  $t$  by stationarity and can be computed from the joint pdf of two successive sampled fading phases

$$P_{q_i, q_j} = \Pr(Q_{t+1} = q_j | Q_t = q_i) \quad (5)$$

$$= \frac{\int_{q_i - \pi/K}^{q_i + \pi/K} \int_{q_j - \pi/K}^{q_j + \pi/K} p(\phi_t^a, \phi_{t+1}^a) d\phi_t^a d\phi_{t+1}^a}{\int_{q_i - \pi/K}^{q_i + \pi/K} p(\phi_t^a) d\phi_t^a} \quad (6)$$

where the marginal pdf is uniform, and the joint pdf is [18]:

$$p(\phi_t^a, \phi_{t+1}^a) = \frac{1 - \rho^2}{4\pi^2} \left[ \frac{\sqrt{1 - B^2} + B(\pi - B \cos^{-1}(B))}{(1 - B^2)^{3/2}} \right]$$

where  $B = \rho \cdot \cos(\phi_{t+1}^a - \phi_t^a)$  and  $\rho = \mathcal{J}_0(2\pi f_D T)$ .

The model described above is essentially an approximation in a dual sense: First, it maps all real fading angles  $\phi^a \in [-\pi, \pi)$  to a finite number of “quantized fading phase states”  $q_i$ ,  $i = 0, 1, \dots, K-1$ . Moreover, the model approximates the dynamics of the continuous process  $\{\phi_t^a\}_{t=0,1,\dots,\infty}$  with a discrete Markov chain, taking values in the finite-state space  $\mathcal{Q}$  and having stationary probabilities  $p_i = 1/K$  and transition probabilities  $P_{q_i, q_j}$ . It should be noted, however, that the channel estimation algorithm in Rayleigh fading does not perceive the channel phase as quantized to the  $K$  values of (4); the algorithms of Section III remain aware of the continuous nature of  $\phi_t^a$  but are based on the assumption that  $\phi_t^a$  is uniformly distributed in a sector of length  $2\pi/K$ , given that it belongs in that sector. Hence, the finite-state Markov model is merely a way to

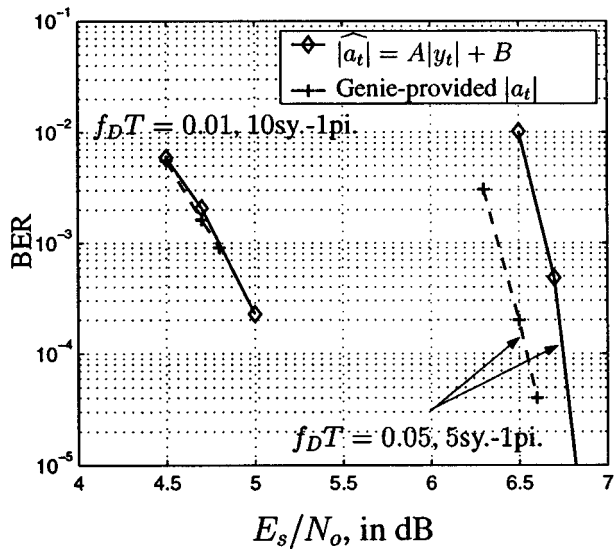


Fig. 1. BER of turbo code in different  $f_D T$  and for different pilot insertion ratios. Performance with affine fading amplitude estimation from the received amplitude is very close to the case when the receiver has access to genie-provided channel amplitude  $|a_t|$ .

reduce the infinite cardinality of the set of possible phases  $\phi_t^a$  to  $K$  “phase states,” such that reliabilities can be assigned to them by the Forward–Backward algorithm.

A Markov model closely related to the one described above was derived in [14] to model the amplitude fading of a phase-coherent Rayleigh fading channel. Here we recognize the phase distortion in a Rayleigh channel as a more severe problem for PSK transmission than amplitude fading. This qualitative observation, along with the difficulty of obtaining coherent phase reference in a high-Doppler low-SNR environment, indicates that the main channel estimation effort for PSK turbo codes should be devoted to acquiring phase coherence rather than exact estimates for the fading amplitude  $|a_t|$ . Thus, the two receivers derived in the next section create the finite-state Markov (FSM) model outlined above and use it to estimate the channel phase with the Forward–Backward algorithm [7]. For the fading amplitude estimation they rely on a simple MMSE symbol-by-symbol affine estimator from the received amplitude of the form  $\widehat{|a_t|} = A|y_t| + B$ . As shown in Fig. 1, this method for amplitude estimation, combined with Forward–Backward phase estimation on the FSM phase model, performs only slightly worse than when having perfect channel amplitude knowledge at the receiver. This result supports the decision to use a simple estimator for the fading amplitude and reserve the power of the Forward–Backward algorithm for phase estimation.

### III. JOINT DATA AND CHANNEL ESTIMATION

Optimal decoding in flat fading (either maximum likelihood sequence detection (MLSD) or minimum bit error rate-maximum *a posteriori* (MAP)-decoding) requires either explicit or implicit estimation of the channel. In the first part of this section we compare iterative phase estimation via the FSM phase model developed above against other channel acquisition methods, such as optimum filtering of pilots. The remaining

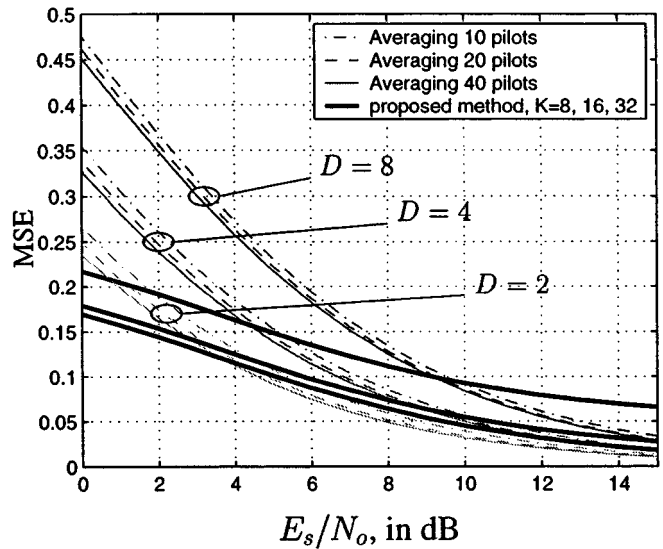


Fig. 2. MSE of channel estimation for a Rayleigh fading channel with  $f_D T = 0.05$  under two scenarios: thin lines represent MSE after Wiener filtering of 10, 20, or 40 pilot symbols spaced one every  $D$  data symbols, while thick lines show the MSE resulting from simple affine amplitude estimation using only the received amplitude and uniform phase uncertainty within a sector of  $2\pi/K$  radians,  $K = 8, 16, 32$ , given that the correct sector is known. Note that this largely corresponds to the estimation procedure followed in the sequel. Observe that for low SNR the second approach is better.

parts of the section describe the FSM-model-based receiver algorithms in detail and present simulation results.

#### A. Quantized Phase Estimation

The performance of turbo codes in flat fading was examined in [19]. In [12] and [13] it was recognized that additional performance benefits are possible when moving from “one-shot” channel estimation (e.g., from pilots) to iterative estimation, integrated with turbo decoding. Specifically, Valenti and Wornner in [13] perform channel estimation via optimum (Wiener) filtering of symbols at each iteration (only pilots at first, and all symbols in subsequent iterations taking into account their reliability), thus exploiting extrinsic information produced by the turbo decoder about coded symbols. In this paper, we also apply the principle of iterative channel estimation, but not with filtering of pilots and coded symbols. Instead, we employ the Forward–Backward algorithm for “quantized phase” estimation based on the FSM model derived in Section II. In Section III-C joint phase estimation and turbo decoding proceed along a supertrellis, constructed by merging the trellises of the code and the Markovian channel state structure. In Section III-D we demonstrate better performance by using a separate Forward–Backward algorithm to estimate the phase state, operating on the trellis of the FSM phase model and exchanging soft information with the constituent decoder SISO [8].

Notice that in both approaches the Forward–Backward algorithm operates on the  $K$ -state Markov phase model and obtains soft phase estimates in the form of a probability distribution on the  $K$  phase states at each time instant (implicitly in the supertrellis, explicitly with the separate trellises). Fig. 2 demonstrates the rationale behind this choice of “quantized phase” estimation: at the low SNR where turbo codes operate, it is advan-

tageous to have a phase estimate with small precision but high reliability (as is possible with the Forward–Backward algorithm on a finite-state phase model) rather than continuous valued estimates of limited reliability, like those provided by optimum filtering, which exhibits higher mean-squared error (MSE).

Under the assumption of knowing the correct phase state (i.e., sector), which is a reasonable assumption with the Forward–Backward algorithm, Fig. 2 indicates that at low SNR and high Doppler (or, equivalently more sparse pilot spacing under the chosen Doppler of  $f_D T = 0.05$ ) the “quantized phase” estimation approach outperforms Wiener filtering solutions in the estimation MSE sense, even with the coarse symbol-by-symbol amplitude estimation  $|\hat{a}_t| = A|y_t| + B$ . In fact, it is this amplitude error that accounts for most of the estimation MSE, as seen from the fact that beyond  $K = 32$  no additional MSE gain is obtained by increasing the number of “quantized phase states.” Of course, at high SNR optimum linear pilot filtering works better than our nonlinear “quantized phase” estimation, even if this is assumed to always identify the correct sector. However, “quantized” estimation seems to have an advantage in the very noisy region of operation of turbo codes, insofar as the channel estimation MSE is an indication of turbo-code bit error rate performance. For a fixed SNR the sequence of thin curves shows decreasing MSE from linear filtering for increasing pilot density. This can be thought of as the improvement to channel estimation expected as the Wiener filter acquires more knowledge about the transmitted data through the turbo iterations. Even in this case, (e.g., when every third symbol becomes perfectly known, i.e.,  $D = 2$ , the lowest thin curve in Fig. 2) if the nonlinear phase estimation guesses the correct sector it promises slightly less MSE at low SNR.

### B. Joint Phase Estimation and Decoding

For “quantized phase” estimation with the Forward–Backward algorithm either on a supertrellis or on separate trellises, the basic quantities needed are the probabilities  $P_{q',q}$  of the channel phase transitioning from state (sector)  $q'$  to  $q$  [given in (5) and (6)] and the likelihoods  $P(y_t|Q_t = q, x_t = x)$ . At time  $t$ , the received complex symbol is  $y_t = |y_t| \cdot e^{j\phi_t^y}$ , where the total received angle  $\phi_t^y$  is the sum of three distinct angles

$$\phi_t^y = \phi_t^x + \phi_t^a + \phi_t^* \quad (7)$$

as shown in Fig. 3. In this figure,  $\phi_t^x$  is the transmitted constellation point angle, as the constituent trellis  $M$ -PSK code transitions from state  $c'$  to  $c$ , i.e.,  $x_t(c' \rightarrow c) = 1 \cdot e^{j\phi_t^x}$ . The fading angle  $\phi_t^a$  is defined from the fading scale factor  $a_t = |a_t| \cdot e^{j\phi_t^a}$ , and  $\phi_t^*$  is the noise-induced additional angle, having probability density function (pdf)  $P(\phi^*)$

$$P(\phi^*; \lambda) = \frac{e^{-\lambda^2}}{2\pi} \left[ 1 + \sqrt{\pi} \lambda \cos \phi^* e^{(\lambda \cos \phi^*)^2} \operatorname{erfc}(-\lambda \cos \phi^*) \right] \quad (8)$$

where  $\operatorname{erfc}(\cdot)$  is the complementary error function and the parameter  $\lambda$  of the distribution depends on the fading amplitude:  $\lambda = |a|/\sigma\sqrt{2}$ . Since the true fading amplitude is unknown, the receiver uses the symbol-by-symbol MMSE affine estimator

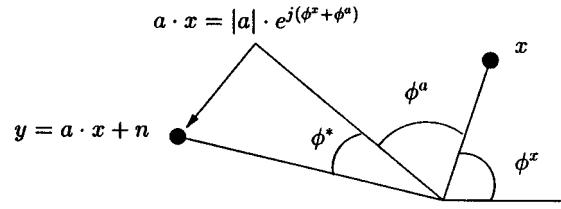


Fig. 3. Addition of angles in fading.

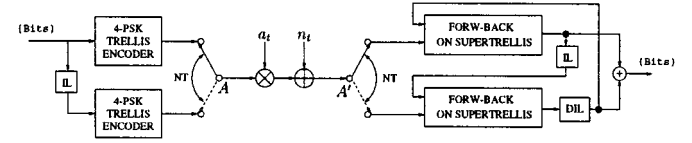


Fig. 4. Block diagram of system employing iterative decoder.

$|\hat{a}_t| = A|y_t| + B$  (with coefficients  $A$  and  $B$  depending only on the SNR of operation). Then it is straightforward to compute the desired likelihood

$$\Pr(y_t|Q_t = q, x_t = x(c' \rightarrow c)) = \Pr(\phi_t^y = \theta|\Pi(\phi_t^a) = q, \lambda_t = \lambda, \phi_t^x = \angle x(c' \rightarrow c)) \quad (9)$$

$$\stackrel{\text{def}}{=} f(\theta|q, \lambda, x) \quad (10)$$

$$= \Pr(\phi_t^x + \phi_t^a + \phi_t^* = \theta|\Pi(\phi_t^a) = q, \lambda, \phi_t^x = \angle x) \quad (11)$$

$$= \Pr(\phi_t^a + \phi_t^* = \theta - \angle x | \phi_t^a \sim \mathcal{U}\left[q - \frac{\pi}{K}, q + \frac{\pi}{K}\right])$$

$$= \frac{K}{2\pi} \cdot \int_{\theta - \angle x - q - (\pi/K)}^{\theta - \angle x - q + (\pi/K)} P(\phi^*; \lambda) d\phi^* \quad (12)$$

where  $P(\phi^*; \lambda)$  was given in (8) and  $\lambda$  is approximated by

$$\hat{\lambda}_t = \frac{A|y_t| + B}{\sigma\sqrt{2}}. \quad (13)$$

Thus, using  $P_{q',q}$  from (6) and  $f(\theta|q, \lambda, x)$  from (12) we proceed to derive the joint algorithms on the supertrellis and on the separate trellises for the code and the channel.

### C. Supertrellis Algorithm

An initial approach to joint estimation and decoding is to combine the Markov model for the quantized fading phase discussed in Section II with the trellis describing the code, to form a *supertrellis*. In essence, the receiver observes the output of a finite-state machine (i.e., the encoder output  $x_t$ ) multiplied with the output of a Markov process (i.e., the “fading phase state”  $Q_t$ ) under AWGN. At time  $t$ , the state  $S_t$  of the supertrellis is an ordered pair consisting of the channel state  $Q_t$  and the code state  $C_t$ , giving  $S_t = (Q_t, C_t) = (q, c) = m$ , with  $m = 0, 1, \dots, 2^\nu K - 1$ , for a code with  $\nu$  memory elements.

Fig. 4 shows the block diagram of the turbo-coded system. Each constituent encoder at the transmitter produces  $\log_2 M$  bits, mapped onto an  $M$ -PSK constellation. The symbols are transmitted into the fading channel in blocks of  $N$ , to preserve the fading phase correlation for the receiver supertrellises. Thus, the switches  $A$  and  $A'$  flip every  $NT$  seconds, where  $N$  is the turbo-code blocklength and  $T$  the baud period.

The receiver consists of two identical modules that run the Forward-Backward algorithm and exchange soft information about the data through a uniform interleaver-deinterleaver pair. Each of them performs *joint* data and channel estimation internally but provides the other estimator with extrinsic information only for the data, since the two blocks of constellation points (pertaining to unscrambled and scrambled data) are transmitted successively into the channel and undergo independent fading. Thus, information about the channel produced by one of the estimators would be irrelevant to the other. However, within each block of  $N$  symbols, the channel is correlated, which facilitates the joint estimation of channel phase and data.

For iterative decoding [8], the crucial quantity to be computed in each supertrellis is

$$\gamma_t(m', m) = \Pr(y_t, S_t = (q, c) | S_{t-1} = (q', c')) \quad (14)$$

$$= \Pr(S_t = (q, c) | S_{t-1} = (q', c')) \cdot \Pr(y_t | S_{t-1} = (q', c'), S_t = (q, c)). \quad (15)$$

For the first term of (15) we have

$$\begin{aligned} \Pr(S_t = (q, c) | S_{t-1} = (q', c')) \\ = \Pr(\Pi(\phi_t^q) = q | \Pi(\phi_{t-1}^{q'}) = q') \\ \cdot \Pr(u_t \text{ such that } C_t = c | C_{t-1} = c') \end{aligned} \quad (16)$$

$$= P_{q', q} \cdot P(u_t; I) \quad (17)$$

where  $P(u_t; I)$  denotes the extrinsic information about the input  $u_t$  provided by the other soft decoder. The second term of (15) is clearly  $f(\theta|q, \lambda, x)$  as defined in (10)–(12).

Note that the algorithm described above can be used with or without pilot symbols. The transition metric  $\gamma_t(m', m)$  of (14) connects only superstates  $(m', m)$  with valid code state transitions ( $c' \rightarrow c$ ). In the case of pilots injected in the coded data stream, the code state does not change, and the only valid supertrellis branches are those with  $c = c'$ . Here we only present simulation results with no pilot symbols.

Fig. 5 presents the simulated BER performance of the system depicted in Fig. 4 under Rayleigh fading with  $f_D T = 0.05$ . The constituent codes are identical, eight-state, recursive systematic rate-1/2, Gray-labeled 4-PSK codes, with maximum effective Hamming distance. They are fully described by the octal parity polynomials  $h_0 = 15$  and  $h_1 = 17$ . The number of quantized phases was  $K = 8$ , resulting in 64-state supertrellises, and the blocklength was  $N = 5000$  symbols. For this relatively high Doppler rate the performance is about 6.5 dB worse than when the same turbo code operates under the ideal assumptions of perfect interleaving and perfect CSI. However, this gap is not very informative, since the constrained capacity of the two channels considered with uniform i.i.d. 4-PSK inputs is quite different at this high Doppler. The vertical dashed line marks the capacity of the idealized scenario of perfectly known  $a_t$  at receiver. It is simply  $I(X; Y|A)$ , a weighted average of the AWGN capacity under the Rayleigh distribution  $p_A(a) = 2ae^{-a^2}$ , giving  $E_b/N_o = -0.08$  dB for the rate 1/2 of interest. The capacity is smaller when CSI is unavailable at the receiver and has to be estimated from received values (much smaller for larger Doppler rates and zero in the limit of i.i.d. fading). A more detailed discussion about constrained 4-PSK capacity under fading follows

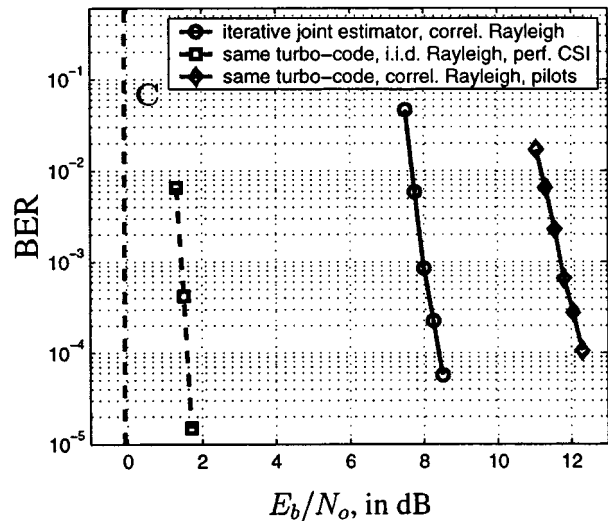


Fig. 5. Supertrellis and noniterative pilot filtering performance in Clarke's channel with  $f_D T = 0.05$ . The dashed curve shows performance of the same turbo code in the same channel with ideal interleaving and perfect CSI at the receiver. The dashed vertical line marks the capacity in this ideal case.

in Section IV-B. To demonstrate the difficulty of obtaining accurate CSI in a practical system at high Doppler, we also simulated a pilot-symbol assisted system with the same turbo code. Specifically, a more sophisticated variant of pilot averaging in [20], using three pilot symbols every five data symbols performs almost 4 dB worse than our joint iterative estimator with no pilot symbols at all. Even if we plot against  $E_s/N_o$  disregarding the sacrifice of  $3/8 = 37.5\%$  in the rate of the pilot system [20], the supertrellis system is still almost 2 dB better. The reason is that essentially every coded symbol with the supertrellis iterations becomes somewhat a pilot, as its reliability increases.

The supertrellis receiver designed and simulated in this section has advantages and limitations. An obvious advantage is its ability to work without external acquisition circuitry or pilot symbols at relatively high Doppler rate. The low rate of each constituent code (here 1/2) compensates for the absence of pilot symbols, allowing the supertrellis algorithm to determine whether a change in the received phase is due to the code or to a change in the channel. Thus, although this scheme does not lose rate directly because of pilots that bear no information, it is the rate reduction inherent in the constituent encoder design that makes channel estimation possible. On a higher level this can be viewed as incorporating the training in the code design, instead of explicitly injecting pilot symbols in the coded data stream of a higher rate code.

The main limitation is computational complexity, since the number of states in each supertrellis is the product of the code states and the number of phase intervals  $K$ . If  $M$ -PSK is used, then  $K \geq 2 \cdot M$  for reasonable phase estimation. This leads to at least 64-state supertrellises with 4-PSK and 128-states with 8-PSK for 8-state constituent codes. Another limitation concerns diversity. The channel estimation procedure along the supertrellis precludes channel interleaving, because the algorithm relies on the correlation between successive phases. Hence, only implicit diversity, due to the interleaver between constituent codes, is provided.

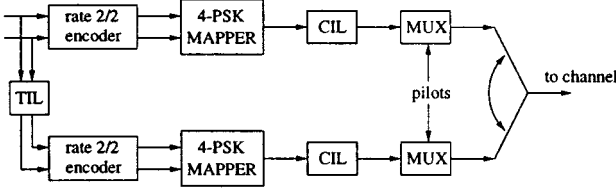


Fig. 6. Transmitter block diagram for pilot-aided turbo code.

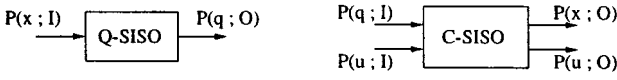


Fig. 7. Basic SISO building blocks of the receiver.

#### D. Algorithm on Separate Trellises

In this section we derive and simulate a better structure for joint channel estimation and turbo decoding based on the Forward–Backward algorithm running on separate trellises for the channel phase and the code. Fig. 6 shows the transmitter block diagram, where the constituent encoders are the best eight-state, rate-2/2 code fragments [21], [22], each producing one systematic and one parity bit per 2-bit input, and their outputs are mapped onto a Gray-labeled 4-PSK constellation. Notice the difference between the turbo interleaver (TIL) and the channel interleaver (CIL), which can be a regular block interleaver. Pilots are injected into the coded data stream at a rate of  $Z \geq 1$  pilots every  $D$  coded PSK symbols, and the blocklength is  $N = 4100$ . Thus, for every  $2N$  input bits, a total of  $2N(D + Z/D)$  symbols are transmitted in flat Rayleigh fading.

Fig. 8 shows the receiver block diagram expanded in the direction of processing time, only to show the potential for a parallel or pipelined implementation. Each of the two main building blocks, denoted **Q** and **C**, implements the Forward–Backward algorithm (Q-SISO and C-SISO) on a separate trellis that describes the Markov channel phase and the code, respectively, exploiting extrinsic information taken from the other block after the appropriate interleaving/deinterleaving operation—T(D)IL or C(D)IL. These two SISO modules and their input/output functionality are depicted in Fig. 7. In the customary SISO notation of [8]  $P(\cdot; I)$  is extrinsic information about a certain quantity (the channel phase state  $q$  or the input symbol  $u$  or the output symbol  $x$ ) entering the SISO block, while  $P(\cdot; O)$  is the updated extrinsic information about the pertinent quantity (again  $q$ ,  $u$ , or  $x$ ) at the output of the corresponding Forward–Backward (SISO) block.

The states are quantized channel phases  $q$  for Q-SISO and code states  $c$  for C-SISO, and the crucial quantity to compute is  $\gamma_t(\text{state}_{t-1}, \text{state}_t)$ . Specifically, for Q-SISO

$$\begin{aligned}
 \gamma_t(q', q) &= \Pr(y_t, Q_t = q | Q_{t-1} = q') \\
 &= \sum_x \Pr(y_t, Q_t = q, x_t = x | Q_{t-1} = q') \\
 &= P_{q', q} \cdot \sum_x \Pr(x_t = x) \cdot \Pr(y_t | Q_t = q, x_t = x) \\
 &= P_{q', q} \cdot \sum_x P(x; I) \cdot f(\theta | q, \lambda, x) \quad (18)
 \end{aligned}$$

where the phase state transition probability  $P_{q', q}$  is precomputed for known Doppler via (5) and (6), and  $\theta$  is the received angle. The function  $f(\theta | q, \lambda, x)$  is defined as in (10). Note that the Q-SISO operates on the whole received block of symbols, coded and pilots alike, but only outputs  $P(q; O)$  for the coded symbols, because information for the channel state during a pilot transmission is irrelevant to the C-SISO. Thus, at a time when a pilot is processed, the above summation is trivial (only one possible  $x$  has nonzero probability) and  $P(q; O)$  is not produced; just the  $\alpha(q)$  and  $\beta(q)$  quantities are updated in the channel trellis. For the C-SISO, we proceed with a similar computation, simpler in this case, since only coded data are processed and there is only one operating mode

$$\begin{aligned}
 \gamma_t(c', c) &= \Pr(y_t, C_t = c | C_{t-1} = c') \\
 &= \sum_q \Pr(y_t, C_t = c, Q_t = q | C_{t-1} = c') \\
 &= \Pr(u | c' \rightarrow c) \\
 &\quad \cdot \sum_q \left\{ \Pr(Q_t = q) \right. \\
 &\quad \quad \cdot \left. \Pr(\phi_t^y = \theta | Q_t = q, x_t = \mathcal{L}x(c' \rightarrow c)) \right\} \\
 &= P(u; I) \cdot \sum_q P(q; I) \cdot f(\theta | q, \lambda, x(c' \rightarrow c)). \quad (19)
 \end{aligned}$$

Again, here  $P(u; I)$  is the extrinsic about the information symbol (2 bits) passed by the other C-SISO, while  $P(q; I)$  is the extrinsic information about the channel state, provided by the Q-SISO.

The performance of the proposed receiver has been checked in two flat fading channels, with Doppler rates 0.01 and 0.05, and the results are shown in Figs. 9 and 10, respectively, for various pilot insertion rates. For comparison purposes, the plots include the performance of the same turbo code when perfect CSI is available at the receiver. The vertical lines show the SNR where capacity equals the transmitted rate, assuming perfect CSI at the receiver (which is a lower bound to the SNR where capacity equals the transmitted rate when realistic channel estimation is performed through pilots). If not explicitly stated otherwise, the simulation was performed by considering  $K = 8$  channel states at the receiver. From the two cases shown, setting  $K = 16$  offers no improvement for  $f_D T = 0.01$ , but some improvement for  $f_D T = 0.05$ . The solid performance curves with different pilot spacing  $D$  provide some insight into the question of what pilot density is required for a given channel dynamic, as this is expressed by the Doppler rate  $f_D T$ .

Note that in the slower changing channel ( $f_D T = 0.01$ ) the performance is about 1.5 dB away from the case when perfect CSI is available. The performance difference from perfect CSI is much more pronounced (about 4.5 dB) in Fig. 10, because the channel is less strongly correlated when  $f_D T = 0.05$ , which makes the estimation task more difficult, so the corresponding SNR (or capacity) penalty resulting from the lack of perfect channel estimates is larger. In general it is difficult to

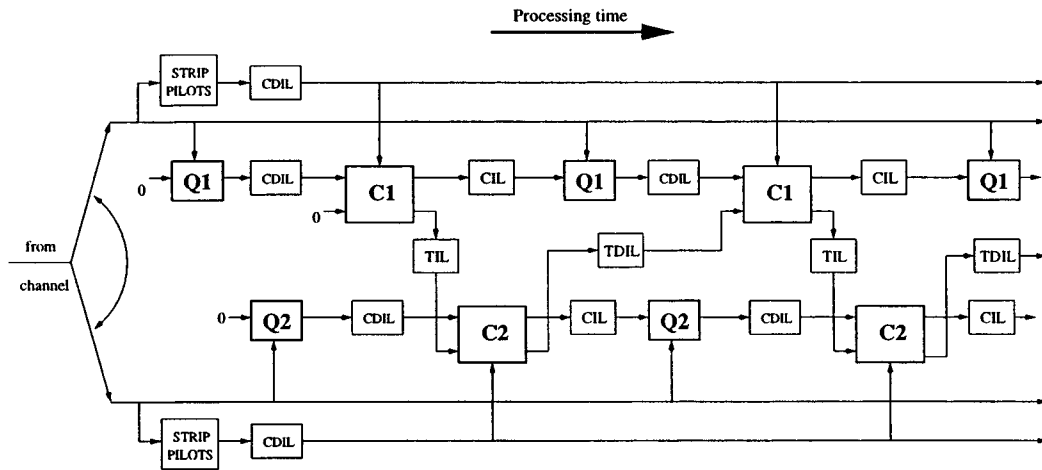


Fig. 8. Receiver expansion in the processing time axis. Notice the possibility of parallel or pipelined operation.

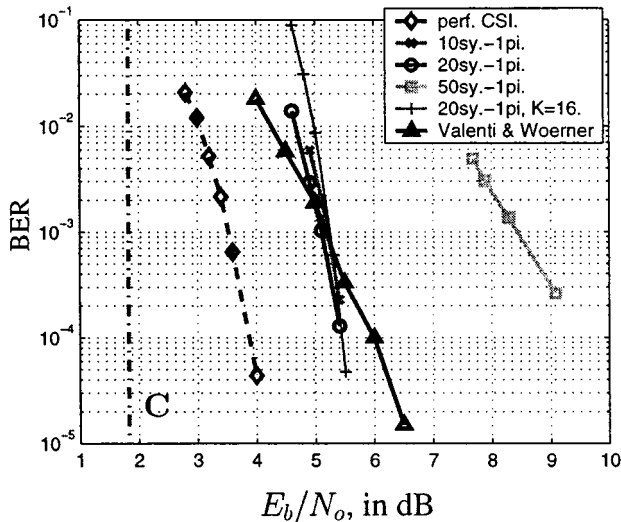


Fig. 9. BER of turbo code, for  $f_D T = 0.01$  and for different pilot insertion rates. Performance lags behind the perfect CSI case by about 1.5 dB. For perspective, the curve marked with triangles is the performance reported by Valenti and Woerner, with a similar turbo-coded system with half the spectral efficiency as the one presented here.

determine quantitatively exactly how the channel capacity is affected by the channel dynamics in order to quantify the intuitive statement made above. A detailed discussion on the effect of the rate of change of the channel on the capacity for a simplified purely Markovian channel model related to Clarke's flat Rayleigh fading is the topic of the next section. Nevertheless, observe that a small number of pilots permits us to increase the overall rate of the system with separate trellises to 1 bit/s/Hz (excluding the pilots) relative to the rate of 1/2 bit/s/Hz for the supertrellis receiver. Furthermore, the complexity of the separate trellises approach is much smaller (the Q-SISO has  $K = 8$  or 16 states and the C-SISO eight states), and the BER performance improves. However, these positive impacts on the rate, complexity, and performance come at the expense of larger latency, due to channel interleaving.

Specifically, the two methods for joint channel estimation and turbo decoding discussed in this section demand the following

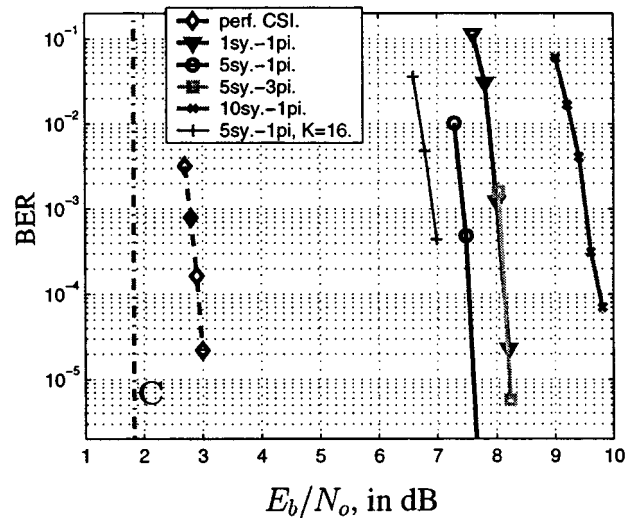


Fig. 10. BER of turbo code, for  $f_D T = 0.05$  and for different pilot insertion rates.

in terms of complexity and latency, for every decoder iteration (in parentheses the numbers in our 4-PSK simulations):

- Supertrellis: two Forward-Backward algorithms, each on a supertrellis with  $2^\nu \cdot K$  states (64 states) and  $2K$  branches (16 here) emanating from each state. This number can be reduced by pruning less likely transitions to phase states far apart from the current one, particularly for small Doppler rates. This receiver suffers no extra latency from channel interleaving.
- Separate trellises: less complexity, four Forward-Backward algorithms in all. Two for the two fully connected  $K$ -state Q-SISOs ( $K = 8$ ), and two for the  $2^\nu$ -state C-SISOs ( $\nu = 3$ ); higher latency because of channel interleaving at the transmitter and de-interleaving at the receiver. Here we implemented those block interleavers to be of equal size to the turbo-code blocklength  $N$ , but this is not necessary, particularly for high Doppler rates. The turbo-code latency is obviously unavoidable.

Finally, for comparison, the complexity of a pilot-averaging system obtaining one-shot channel estimates without iteration

is just two  $2^{\nu}$ -state C-SISOs per iteration, plus the additional one-time cost of pilot filtering per block. The latency from channel interleaving is the same as in the separate trellises scenario. Thus, one-shot Wiener filtering of pilots has much less complexity but misses the benefit of interaction of the estimation and the decoding procedures, which assist each other in our joint estimation schemes.

#### IV. CHANNEL CAPACITY

##### A. Simplified Finite-State Markov Channel (FSMC) Model

For Clarke's flat Rayleigh channel of (1), where the process  $\{a_t\}$  is stationary and ergodic, the definition of the capacity in Gallager [23] applies

$$C = \lim_{N \rightarrow \infty} \frac{1}{N} \cdot I(X^N; Y^N) \quad (20)$$

where  $X^N$  and  $Y^N$  denote sequences of channel inputs and outputs, respectively. Here we are interested in the constrained capacity for inputs from a finite uniform constellation, such as the 4-PSK we use. If the decoding delay is constrained to be small enough relative to the decorrelation time of the channel, then no positive rate is achievable, and outage probability, not capacity, becomes the correct performance measure [24]. In this discussion decoding delay will not be constrained, hence the capacity definition (20) is valid. However, direct computation of the capacity of Clarke's channel (1), with the process  $\{a_t\}$  having autocorrelation and power spectral density given by (2) and (3), is an open problem, although work has been done considering a piecewise constant channel or exponential autocorrelation; see [25]. Another body of work has determined the capacity of any finite-state Markovian channels (FSMC) in [4] and more generally in [5]. This is useful, because in practice an FSMC becomes a good model that mimics any channel statistics, if the number of states is chosen large enough. Here, we study the following FSMC—admittedly more benign than Clarke's, but largely equivalent and capturing most of the performance-driving phenomena of (1)

$$y_t = e^{jQ_t} \cdot x_t + n_t, \quad t = 0, 1, 2, \dots \quad (21)$$

where  $\{Q_t\}_{t=0,1,\dots}$  is the discrete  $K$ -state Markov chain of Section II, taking the values (4). In other words, the FSMC of (21) induces no amplitude fading but rotates the transmitted phase by a discrete amount, correlated in time according to the Markov model of Section II, and adds AWGN.

The channel models of (1) and (21) have differences and similarities. For instance, the process  $\{Q_t\}$  is not strictly bandlimited, unlike Clarke's  $\{a_t\}$  process. Moreover, the FSMC only adds discrete phase distortion and no amplitude fading, hence it is more benign. Despite those differences, results obtained for the FSMC largely carry over to the more realistic channel model of Clarke, uniformly shifted by about 1–1.5 dB. The relative ordering of simulations remains unchanged. This, along with the mathematical tractability of the FSMC in terms of capacity bounds, is the reason we focus attention on the FSMC model of (21) in this section.

Note that to simulate in this FSMC, the only thing that has to be modified in the algorithms derived so far is the definition of  $f(\theta|q, \lambda, x)$  of (10). Since the channel phase is now discrete,

there is no need for integration as in (12) and  $\lambda$  is fixed. So for the FSMC

$$f(\theta|q, \lambda, x) = P\left(\phi^* = \theta - q - \angle x; \lambda = \frac{1}{\sigma\sqrt{2}}\right) \quad (22)$$

with  $P(\phi^*; \lambda)$  as given in (8). Simulating in the pure FSMC of (21) eliminates the (small) discrepancy of the Rayleigh simulation, in which the channel in fact follows Clarke's model, but the receiver models it as having Markovian transitions between phase sectors. Results from simulations in the FSMC are shown in Section IV-C, after deriving the capacity bounds in Section IV-B.

##### B. Bounds on the Capacity of an FSMC

For the capacity of any stationary ergodic FSMC the definition (20) still applies. The algorithm in [5] (generalizing the results of [4]) offers a way to compute the capacity of a FSMC like the one considered here. However, the computation needed to obtain the limiting distributions of vectors  $\pi_n$  and  $\rho_n$ , where  $\pi_n(k) = p(Q_n = q_k | x^{n-1}, y^{n-1})$  and  $\rho_n = p(Q_n = q_k | y^{n-1})$ ,  $k = 0, 1, \dots, K-1$  (see [5]) is infeasible for a number of states  $K$  in the Markov model beyond  $K = 2$  or 4. However, this is insufficient for our purposes here, since  $K \geq 8$  for 4-PSK is needed to maintain reasonable similarity between the real world fading of Clarke and the FSMC of (21).

A computationally much simpler solution is to upper-bound the constrained capacity of the FSMC described in the previous section. An obvious, easy to compute, but very loose upper bound of  $C_{\text{FSMC}}$  of this Markovian quantized phase-distortion channel is the constrained capacity given the current state  $Q$  of the channel

$$C_{\text{FSMC}} = \lim_{N \rightarrow \infty} \frac{1}{N} \cdot I(X^N; Y^N) \leq \lim_{n \rightarrow \infty} I(X_n; Y_n | Q_n). \quad (23)$$

Clearly this is the constrained capacity of the AWGN channel with PSK inputs, depicted in Fig. 11 as  $I_{UB}(D = 0)$  versus symbol SNR,  $E_s/N_o$ . But it is possible to compute a sequence of progressively tighter upper bounds on  $C_{\text{FSMC}}$ , from the following theorem.

*Theorem 1:* For any finite-state Markov channel (FSMC) with states  $Q$ , a sequence of progressively tighter and asymptotically tight upper bounds to the capacity  $C_{\text{FSMC}}$  is

$$I_{UB}(D) = \frac{1}{D} \cdot I(X_1^D; Y_1^D | Q_0, Q_{D+1}), \quad D = 0, 1, 2, \dots, \infty \quad (24)$$

where, for  $D = 0$ , we define the upper bound to be given by (23), or  $I_{UB}(0) = I(X; Y | Q)$ .  $\square$

The proof of this theorem is relegated to the Appendix. From the proof it also becomes obvious how to obtain asymptotically tight lower bounds on  $C_{\text{FSMC}}$ , namely the quantities  $I_{LB}(D) = I_{UB}(D) - (2 \log K/D)$ , where  $K$  is the number of states of the Markov channel. Unfortunately, those bounds are not tight enough to be useful for the first few terms (i.e., for  $D = 0, 1, 2$ ) that we evaluate numerically below.

In Fig. 11 we plot the loose upper bound  $I_{UB}(0)$  of (23), as well as the tighter bounds  $I_{UB}(1)$  and  $I_{UB}(2)$  against the channel SNR, for three FSMCs, with constant unit amplitude and  $K = 8$  phase states, derived from Rayleigh channels with

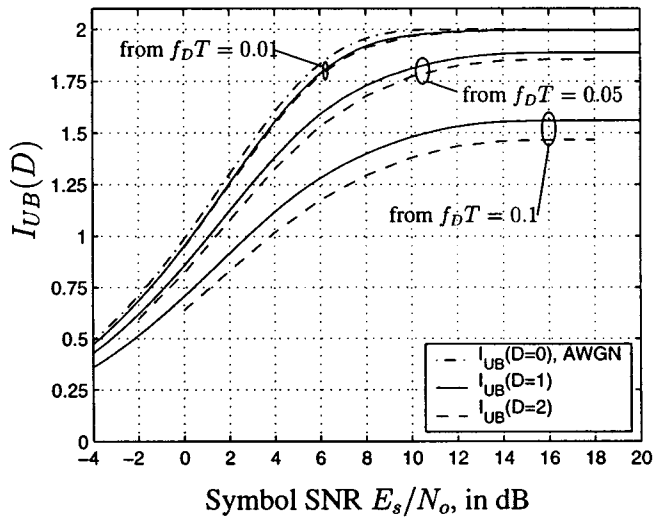


Fig. 11. Bounds on  $C_{\text{FSMC}}$  for i.i.d. 4-PSK inputs, for three Markov-phase channels derived from Rayleigh channels with Doppler rates  $f_D T$ , as described in Section II.

Doppler rates of 0.01, 0.05, and 0.1. Observe the capacity reduction with increasing Doppler rates, which demonstrates the increasing difficulty of reliable channel estimation in faster varying channels, even in cases where the noise is negligible. In the limit of the uncorrelated channel (i.i.d. discrete distortion phases) channel estimation is impossible, and the constrained PSK capacity is identically zero regardless of the SNR. Conversely, when the Doppler rate is 0.01, the bounds show small capacity losses with respect to perfect channel information ( $I_{UB}(0)$ ), indicating that in this case very good channel estimates can be obtained due to the strong time correlation.

From a different viewpoint the bounds  $I_{UB}(D)$ ,  $D > 0$  are not just capacity bounds in a case with no pilot symbols available at the receiver.  $I_{UB}(D)$  also upperbounds capacity for a pilot-aided system, whereby groups of  $Z \geq 1$  pilot symbols are injected into the coded data stream, spaced  $D$  or more coded symbols apart. This is true because  $I_{UB}(D)$  decreases in  $D$ , and no group of pilots can ever offer better estimates of the channel for  $t = 1, \dots, D$  than noiseless knowledge of the channel states  $Q_0, Q_{D+1}$  at the outer edges of each group of  $D$  information symbols.

### C. Performance in the FSMC Relative to Capacity

This section presents the BER performance of the algorithms discussed in Section III, simulated in the FSMC of (21). The only adaptation needed in the joint phase estimation and turbo decoding algorithms for the supertrellis as well as the separate trellises is shown in (22).

Fig. 12 shows the supertrellis simulation result. Note that both the solid curve (supertrellis in FSMC) and the dashed curve (same turbo code and channel with perfect interleaving and CSI at the receiver) are about 1 dB better than their Rayleigh counterparts in Fig. 5. Also, the solid vertical line marks the  $E_b/N_o = 0.2$  dB of the capacity bound  $I_{UB}(2)$ , at rate 1/2 bits/Hz. This shows that the supertrellis receiver, due to the absence of pilots and channel interleaving performs quite far from the tightest capacity bound. In contrast, the same turbo code with perfect inter-

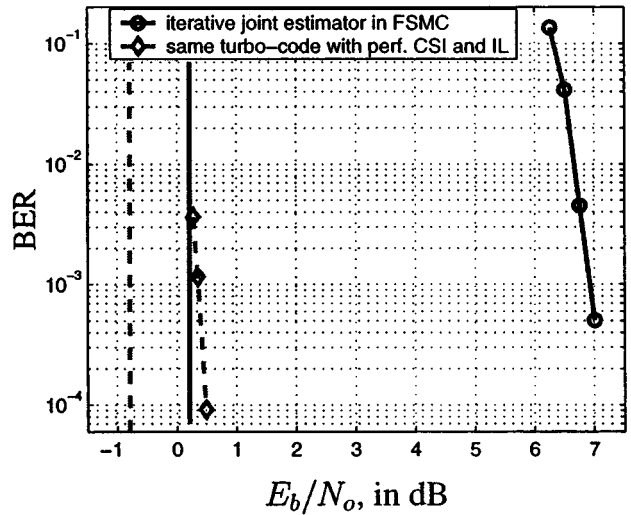


Fig. 12. BER versus  $E_b/N_o$  for the turbo code in the FSMC with  $K = 8$  quantized phases derived from  $f_D T = 0.05$ . The solid vertical line shows the capacity bound  $I_{UB}(2)$ , while the dashed curve and line show performance and capacity respectively given perfect interleaving (IL) and CSI.

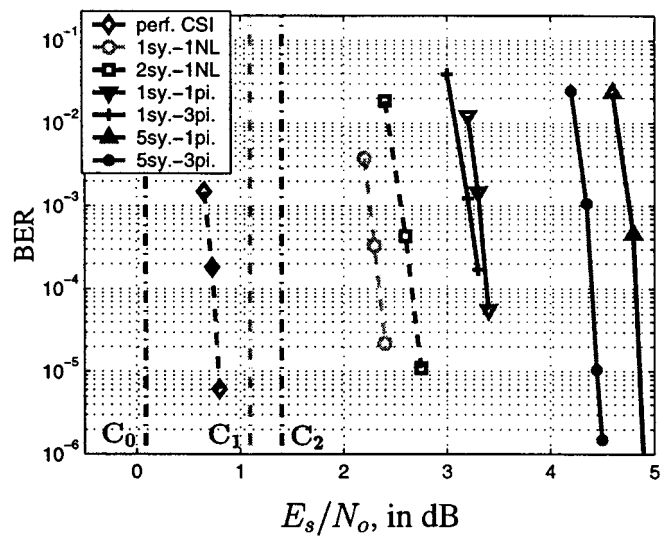


Fig. 13. Capacity bounds and simulated BER in Markov channel for various pilot insertion rates. Vertical lines labeled  $C_0, C_1, C_2$  mark the  $E_s/N_o$  where the capacity bounds  $I_{UB}(D)$ ,  $D = 0, 1, 2$  respectively reach rate 1. Simulations with noiseless pilots (exact channel knowledge) every  $D = 0, 1, 2$  coded symbols, for which  $C_0, C_1, C_2$  are exact capacities are plotted with dashed curves, marked with diamonds, circles, and squares, respectively. Solid curves are simulations with usual noisy pilots. SNR is plotted as  $E_s/N_o$  to keep distance from capacity bounds consistent.

leaving and CSI performs very close to capacity, which is given by  $I_{UB}(0)$  and marked by the vertical dashed line in Fig. 12.

In Fig. 13, we show simulation results of the joint estimator on separate trellises in the Markov phase channel with  $K = 8$  discrete phases, derived from  $f_D T = 0.05$ . The conventional simulations (solid curves) are with  $D = 5$ , and  $D = 1$ , namely injecting pilot symbols every five or every other coded data symbol. The dashed curves show simulation results when injecting one noiseless pilot, i.e., exact CSI, every  $D = 1$  or 2 symbols. The dashed curve marked with “o” shows performance with perfect CSI everywhere. All three of these simulation scenarios are idealized, since no noiseless pilots or perfect CSI

TABLE I  
SIMULATION RESULTS FOR THE PILOT-AIDED TURBO CODE IN THE FSMC  
DERIVED FROM  $f_D T = 0.05$

$D$	$Z$	pilot nature	SNR @ $10^{-4}$	SNR Gap From	
				CSI	$C_{1,2}$
1	1	noisy	3.38 dB	2.58 dB	2.28 dB
1	3	noisy	3.26 dB	2.46 dB	2.16 dB
1	1	noiseless	2.4 dB	1.6 dB	1.3 dB
2	1	noiseless	2.7 dB	1.9 dB	1.3 dB
5	1	noisy	4.86 dB	4.06 dB	3.46 dB
5	3	noisy	4.4 dB	3.6 dB	3.0 dB
5	1	noiseless	3.4 dB	2.6 dB	2.0 dB
10	1	noisy	5.95 dB	5.15 dB	4.55 dB
10	3	noisy	5.51 dB	4.71 dB	4.11 dB
10	1	noiseless	4.81 dB	4.01 dB	3.41 dB

can be available, but they are cases where the capacity bounds  $I_{UB}(D)$ ,  $D = 0, 1, 2$  are no longer a bound, but the true capacity of these idealized transmission situations. Observe that the distance of the “noiseless” pilot simulated BER curves for  $D = 1$  and 2 from the respective capacity bounds (dashed vertical lines marked  $C_1, C_2$ ) is the same, a consistent 1.3 dB. Also note that performance is plotted against symbol SNR  $E_s/N_o$ , unlike previous curves, in order to show the consistent SNR gap with respect to the capacity bounds.

The results of simulations in the Markov phase channel (those included in Fig. 13 and others) are summarized in Table I in the following fashion. The first column ( $D$ ) shows how many consecutive information symbols are transmitted into the channel before  $Z = 1$  or more (in the second column) consecutive pilot symbols are injected. In cases marked “noisy” in the third column, the pilot symbols are conventional pilots, offering noisy estimates of the channel at the edge-points of a  $D$ -symbol information block. Cases marked “noiseless” (NL) refer to the unrealistic scenario of a noiseless pilot, offering *exact* channel knowledge between groups of  $D$  symbols. For those cases the bounds  $I_{UB}(D)$  are no longer bounds, but the exact capacity for this idealized transmission scheme. The fourth column of the table shows the  $E_s/N_o$  in decibels, at which a BER of  $10^{-4}$  is reached. The last two columns show the SNR gap between the simulated performance and performance with perfect CSI everywhere, and the SNR gap from the tightest appropriate capacity bound computed in Section IV-B. Observe that the gap from the capacity bounds is *smaller* than that from perfect CSI, which shows that in this noisy fast-varying channel the performance with perfect CSI is not achievable. For the first three rows, the correct capacity bound is  $C_1 = 1.1$  dB, i.e., the  $E_s/N_o$  where  $I_{UB}(D = 1)$  reaches rate 1. For all other rows, the tightest bound is  $C_2 = 1.4$  dB, i.e., the SNR at which  $I_{UB}(D = 2)$  reaches rate 1.

The results in Table I are not surprising. They indicate steady improvement as the number of pilot symbols and the rate at which pilot groups are injected increase. Obviously the scenarios of the third and fourth row are closest to capacity, and the SNR gap increases as groups of pilots are further apart from each other ( $D$  increases). It must be noted that results in Fig. 13 and Table I are in terms of  $E_s/N_o$  and thus do not take into account the energy expended on pilots, because this makes their

correspondence with the capacity bounds more clear. For instance, if shown against  $E_b/N_o$ , row 6 of Table I would be worse than row 5, because a lot of energy is wasted on evidently unnecessary pilot symbols.

On a related issue, for the case on the first row of Table I with  $Z = 1$  and  $D = 1$ , BER of  $10^{-4}$  is reached at  $E_s/N_o = 3.38$  dB, or  $E_b/N_o = 6.38$  dB, since the information rate is 1/2 bit/symbol, as reduced by the pilots from the rate-1 turbo code. The same BER at the same information rate is achieved at  $E_b/N_o \geq 7.0$  dB with the supertrellis receiver; see Fig. 12, because the turbo code there is of rate 1/2. This performance improvement of at least 0.6 dB highlights the positive effect of channel interleaving in providing diversity for the separate trellises approach, despite the larger complexity of the supertrellis.

## V. CONCLUSION

Two methods to combat flat fading without having access to explicit channel state information (CSI) at the receiver are shown. With both methods, the receiver forms a finite-state Markov model for the fading channel phase. With the first approach, the finite state machines of the channel and the code are combined to form a supertrellis, along which the channel and the data are jointly being estimated, without pilot symbols or channel interleaving. The second method employs separate trellises for decoding and phase estimation. This method uses pilot symbols and channel interleaving with higher rate codes and provides good performance approaching upper bounds to capacity with reasonable complexity. The estimation of channel phase and data is done jointly, on the supertrellis or on the separate trellises, via the Forward-Backward algorithm.

## APPENDIX

This appendix contains the proof of Theorem 1. Although in Section IV we only use the bounds  $I_{UB}(D)$  for the constrained capacity with i.i.d. 4-PSK inputs, these bounds are in fact general and also apply for any input distribution, including the capacity achieving Gaussian. For this reason, in the following proof we use the symbol  $h(\cdot)$  for the entropy of the input variable  $X$ , implying differential entropy, while for  $X$  coming from a 4-PSK constellation the symbol  $H(\cdot)$  for the entropy of a discrete random variable is more appropriate. We first show that the sequence  $I_{UB}(D)$ ,  $D = 0, 1, 2, \dots, \infty$  decreases in  $D$ , or

$$\begin{aligned}
 I_{UB}(D) &\geq I_{UB}(D+1) \\
 &\Leftrightarrow \frac{1}{D} \cdot \left[ h(X_1^D | Q_0, Q_{D+1}) \right. \\
 &\quad \left. - h(X_1^D | Q_0, Y_1^D, Q_{D+1}) \right] \\
 &\geq \frac{1}{D+1} \cdot \left[ h(X_1^{D+1} | Q_0, Q_{D+2}) \right. \\
 &\quad \left. - h(X_1^{D+1} | Q_0, Y_1^{D+1}, Q_{D+2}) \right] \\
 &\Leftrightarrow \frac{1}{D} \cdot h(X_1^D | Q_0, Y_1^D, Q_{D+1}) \\
 &\leq \frac{1}{D+1} \cdot h(X_1^{D+1} | Q_0, Y_1^{D+1}, Q_{D+2}) \quad (25)
 \end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow \frac{1}{D} \cdot \sum_{i=1}^D h(X_i|Q_0, X_1^{i-1}, Y_1^D, Q_{D+1}) \\
&\leq \frac{1}{D+1} \cdot \sum_{i=1}^{D+1} h(X_i|Q_0, X_1^{i-1}, Y_1^{D+1}, Q_{D+2})
\end{aligned} \tag{26}$$

where we start from the definition of mutual information, (25) follows from the fact that the input  $X$  is i.i.d. and independent of all the channel states  $Q$ , while (26) is the chain rule for entropy. Now observe two facts, which hold for  $i = 1, \dots, D$

$$\begin{aligned}
&h(X_i|Q_0, X_1^{i-1}, Y_1^D, Q_{D+1}) \\
&\leq h(X_i|Q_0, X_1^{i-1}, Y_1^{D+1}, Q_{D+2})
\end{aligned} \tag{27}$$

because conditioning reduces entropy, and also

$$\begin{aligned}
&h(X_i|Q_0, X_1^{i-1}, Y_1^D, Q_{D+1}) \\
&= h(X_{i+1}|Q_1, X_2^i, Y_2^{D+1}, Q_{D+2}) \\
&\leq h(X_{i+1}|Q_0, X_1^i, Y_1^{D+1}, Q_{D+2})
\end{aligned} \tag{28}$$

by stationarity and the fact that conditioning reduces entropy. Therefore, if we denote the terms of the sums in the left-hand side and right-hand side of (26)  $l_i, i = 1, \dots, D$  and  $r_i, i = 1, \dots, D+1$  respectively for notational convenience, then (27) implies  $l_i \leq r_i, i = 1, \dots, D$  and (28) implies  $l_i \leq r_{i+1}, i = 1, \dots, D$ . So, assuming  $r_{\max} = r_k$ , we have

$$\begin{aligned}
\frac{1}{D} \cdot \sum_{i=1}^D l_i &= \frac{1}{D+1} \cdot \left[ \frac{1}{D} \sum_{i=1}^D l_i + D \cdot \left( \frac{1}{D} \sum_{i=1}^D l_i \right) \right] \\
&\leq \frac{1}{D+1} \cdot \left[ r_{\max} + D \cdot \left( \frac{1}{D} \sum_{i=1, i \neq k}^{D+1} r_i \right) \right]
\end{aligned} \tag{29}$$

$$= \frac{1}{D+1} \cdot \sum_{i=1}^{D+1} r_i \tag{30}$$

where (29) follows from (27) and (28). This proves (26) and with it the claim that  $I_{UB}(D)$  is decreasing in  $D$ . Also,  $I_{UB}(D)$  is nonnegative, hence it has a limit. We now show that this limit is, in fact,  $C_{\text{FSMC}}$ , or in other words that the bound  $I_{UB}(D), D = 0, 1, 2, \dots$  is asymptotically tight

$$\begin{aligned}
&\frac{1}{D} \cdot I(X_1^D; Y_1^D | Q_0, Q_{D+1}) - \frac{1}{D} \cdot I(X_1^D; Y_1^D) \\
&= \frac{1}{D} \cdot \left[ h(X_1^D | Y_1^D) - h(X_1^D | Y_1^D, Q_0, Q_{D+1}) \right] \\
&= \frac{1}{D} \cdot I(X_1^D; Q_0, Q_{D+1} | Y_1^D) \geq 0
\end{aligned} \tag{31}$$

so  $I_{UB}(D)$  upperbounds every term in the sequence  $1/n \cdot I(X_1^n; Y_1^n), n \geq D$  and hence also upperbounds  $C_{\text{FSMC}}$ , from (20). Finally, it is easy to see that the bounds  $I_{UB}(D), D = 0, 1, 2, \dots$  are asymptotically (in  $D$ ) tight, by examining the difference term

$$\frac{1}{D} \cdot I(X_1^D; Q_0, Q_{D+1} | Y_1^D) \tag{32}$$

$$\begin{aligned}
&= \frac{1}{D} \cdot \left[ H(Q_0, Q_{D+1} | Y_1^D) \right. \\
&\quad \left. - H(Q_0, Q_{D+1} | X_1^D, Y_1^D) \right]
\end{aligned} \tag{33}$$

$$\leq \frac{1}{D} \cdot H(Q_0, Q_{D+1} | Y_1^D) \tag{34}$$

$$\leq \frac{1}{D} \cdot H(Q_0, Q_{D+1}) \tag{35}$$

$$\leq \frac{2 \log K}{D} \tag{36}$$

which converges to zero, as  $D$  grows without bound.

## REFERENCES

- [1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes (1)," in *Proc. ICC '93*, pp. 1064–1070.
- [2] C.-E. W. Sundberg and N. Seshadri, "Coded modulation for fading channels: An overview," *Eur. Trans. Telecom.*, vol. 4, no. 3, pp. 309–323, May–June 1993.
- [3] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–945, May 1998.
- [4] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.
- [5] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 868–886, May 1996.
- [6] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.
- [7] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.
- [8] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Soft-input soft-output modules for the construction and distributed iterative decoding of code networks," *Eur. Trans. Telecom.*, vol. 9, no. 2, pp. 155–172, Mar.–Apr. 1998.
- [9] J. Garcia-Frias and J. D. Villasenor, "Turbo codes for binary Markov channels," in *Proc. ICC '98*, Atlanta, GA, June 7–11, pp. 110–115.
- [10] C. Douillard, M. Jézéquel, C. Berrou, A. Piccart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecom.*, vol. 6, no. 5, pp. 507–511, Sept.–Oct. 1995.
- [11] A. Anastasopoulos and K. M. Chugg, "TCM for frequency-selective, interleaved fading channels using joint diversity combining," in *ICC '98*, Atlanta, GA, June 7–11, pp. 1340–1344.
- [12] H. J. Su and E. Geraniotis, "Improved performance of a PSAM system with iterative filtering and decoding," in *36th Annu. Allerton Conf. Communication, Control and Computing*, Sept. 1998, pp. 156–166.
- [13] M. C. Valenti and B. D. Woerner, "Refined channel estimation for coherent detection of turbo-codes over flat fading channels," *Electron. Lett.*, vol. 34, no. 17, pp. 1648–1649, Aug. 1998.
- [14] H. S. Wang and N. Moayeri, "Finite-state Markov channel: A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 163–171, Feb. 1995.
- [15] W. C. Jakes, Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.
- [16] H. S. Wang and P. C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 353–357, May 1996.
- [17] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, pp. 957–1000, July 1968.
- [18] W. C. Y. Lee, *Mobile Communications Engineering*, New York: McGraw-Hill, 1982.
- [19] E. K. Hall and S. G. Wilson, "Design and analysis of turbo-codes on Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 160–174, Feb. 1998.
- [20] A. G. Dabak, S. Hosur, and T. Schmid, "Iterative channel estimation for wideband CDMA systems," in *36th Annu. Allerton Conf. Communication, Control, and Computing*, Sept. 1998, pp. 441–449.
- [21] C. Fragouli and R. Wesel, "Symbol-interleaved parallel concatenated trellis coded modulation," in *1999 IEEE Communication Theory Mini-conf. conjunction with ICC 99*, Vancouver, BC, Canada, June 6–10.
- [22] —, "Turbo encoder design for symbol-interleaved parallel concatenated trellis-coded modulation," *IEEE Trans. Commun.*, vol. 49, pp. 425–435, Mar. 2001.

- [23] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [24] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 369–378, May 1994.
- [25] G. Kaplan and S. Shamai, "Achievable performance over the correlated Rician channel," *IEEE Trans. Commun.*, vol. 42, pp. 2967–2978, Nov. 1994.

**Christos Komninakis** (S'97) was born in Athens, Greece, in 1972. He received the diploma in electrical and computer engineering from the National Technical University of Athens in 1996 and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles, in 1998 and 2000, respectively.

He is currently with Broadcom Corporation, in El Segundo, CA. His research interests include coding and channel estimation for linear time-varying channels and diversity techniques for increasing the data rate of wireless systems.

**Richard D. Wesel** (S'91–M'96–SM'99) was born in Marietta, OH, in 1966. He received both the S.B. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1989 and the Ph.D. degree in electrical engineering from Stanford University in 1996.

From 1989 to 1991 he was a Member of Technical Staff with AT&T Bell Laboratories. His work at AT&T resulted in two patents. Since 1996 he has been with the University of California, Los Angeles (UCLA) as an Assistant Professor in the Electrical Engineering Department. His research interests include the area of communication theory with particular interest in the topics of channel coding and distributed communication.

Since 1999 Dr. Wesel has been an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of coding and coded modulation. He received a National Science Foundation CAREER Award in 1998. In 1999 he received an Okawa Foundation award. He recently received the 2000 TRW Excellence in Teaching Award from the UCLA Henry Samueli School of Engineering and Applied Science.