Soft-Output Demodulator in Space-Time-Coded Continuous Phase Modulation

Xiaoxia Zhang and Michael P. Fitz, *Member, IEEE*

***Abstract*—Space-time coding has shown great promise for digital transmission in wireless communication links, especially when the channel response is known at the receiver. Space time coding combined with continuous phase modulation (CPM) can offer better tradeoffs in bandwidth and power efficiency. Because of the memory inherent in CPM, channel estimation is often harder than for linear modulations. In this paper, we present an adaptive soft algorithm that performs joint channel estimation and data detection for space-time CPM systems. Properly designed pilot symbols are inserted at the very beginning to give good initial estimates of the channels. This soft receiver is further applied to the interleaved space-time CPM system to yield better performance with moderate complexity through iterative processing. Simulation results show that the receiver can often achieve near-coherent performance in quasistatic fading as well as in time-varying fading.**

***Index Terms*—Channel estimation, continuous phase modulation, iterative decoding, space-time coding.**

I. Introduction

THE major challenge of wireless links arises from multipath induced signal fading. The pioneering work by Telatar [1] and Foschini and Gans [2] shows that multiple antennas in a wireless communication system can dramatically increase the system capacity. For L_t transmit antennas and L_r receive antennas, this work shows that with spatial independence, there are essentially L_tL_r levels of diversity available and that there are $\min(L_t, L_r)$ independent parallel channels that can be established. These information-theoretic studies spawned two lines of work: one where the number of independent channels is small [3], [4] and one where the number of independent channels is large [5]. Space-time (ST) coding [3], [4], [6] is the commonly accepted approach for small number of parallel channels and has been shown to be a very effective diversity strategy. Space-time coding exploits the diversity available in multiple transmit antennas to achieve good performance and bandwidth efficiency. Another power and bandwidth efficient transmission method is continuous phase modulation (CPM) [7], [8]. CPM's constant envelope provides good power efficiency, and the maintained phase continuity can provide good spectral efficiency. When CPM is combined with space-time coding (ST-CPM), better performance can be expected [9].

Manuscript received October 15, 2001; revised May 14, 2002. This work was supported by the National Science Foundation under Grant NCR-9706372. The associate editor coordinating the review of this paper and approving it for publication was Dr. Bertrand M. Hochwald.

X. Zhang is with the Standards Group, Qualcomm, San Diego, CA 92121 USA.

M. P. Fitz is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095-1594 USA.

Publisher Item Identifier 10.1109/TSP.2002.803344.

However, most space-time codes have been proposed and analyzed by assuming that perfect channel state information (CSI) is available at the receiver. Accurate estimation of CSI in space-time coding can be quite challenging since in frequency flat channels, L_tL_r channel gains need to be estimated. Although it has been proven that even in the presence of channel estimation errors, the rank and determinant design criteria are still valid and the space-time codes can still yield improved performance [10], a degradation has to be expected, especially in the time-varying channels.

One way to deal with this problem is to perform joint channel estimation and data detection. This approach has been addressed in [11] and [12], where the expectation–maximization (EM) algorithm [13] is used to estimate the channel. The EM algorithm performs well in quasistatic fading, but its iterative nature makes it inapplicable to the time-varying channels. Recently, based on per-survivor processing (PSP) [14], Viterbi decoding of space-time codes with linear modulation in the absence of channel state information is investigated in [15] and [16]. These PSP-based algorithms provide good tracking performance while releasing hard decisions of the decoded data. The difference between [15] and [16] is that in [15], the LMS algorithm is used to perform channel estimation, whereas decision-directed Kalman filtering is used in [16]. This decision-directed Kalman filtering in space-time coding was previously proposed in [17], where the 2×2 Alamouti space-time block code [18] is considered.

For single transmit antenna case, adaptive soft-output data detection for CPM signals has seen several studies. A demodulator where the channel estimation is done through Wiener prediction filtering [19], [20] has been shown to perform well both in quasistatic fading and in relatively fast fading with single receive antenna or multiple correlated receive antennas. Independent work by Hansson and Aulin [21] presents a similar soft output demodulation algorithm for multiple receive antennas. In this paper, we extend these adaptive soft output demodulation algorithms for multiple transmit antenna systems.

Due to the difficulty of obtaining multiple path gain estimates blindly, we propose including pilot symbols in the frame to give good initial estimates of the channels. The amount of training needed in multiple antenna links with linear modulation is explored by Hassibi and Hochwald in [22]. They show that the smallest number of training symbols that guarantees meaningful estimates equals the number of transmit antennas. How to design pilots with linear modulation has been investigated in [3], [6], and [23], with an orthogonal pilot matrix providing a solution in quasistatic fading. Perfectly orthogonal pilot symbol designs are not usually possible with CPM due to the required phase continuity. Since the pilot symbol design criterion

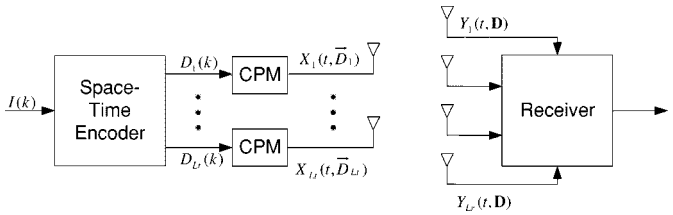

Fig. 1. Block diagram of space-time CPM (ST-CPM).

is different from that in linear modulation, this will also be addressed in this paper. Simulation results demonstrate that with the properly designed pilot symbols and adaptive soft output demodulation, near-coherent performance can be achieved. This adaptive soft-output receiver is further applied to a serial concatenated (SC) system, and iterative detection is performed to gain an improved system performance. The SC-CPM has been investigated in [24] and [25], where the perfect CSI is assumed to be available at the receiver, and neither joint channel estimation and data detection nor transmit diversity is considered.

This paper is organized as follows. Section II describes the system. In Section III, the joint channel estimation and soft-output data detection is derived. Pilot symbol design is investigated in Section IV. In Section V, the derived receiver is applied to conduct the "turbo" processing of the interleaved space-time CPM. Section VI shows the simulation results, and Section VII concludes the paper.

II. Problem Formulation

In this paper, we consider a mobile communication system equipped with L_t transmit antennas and L_r receive antennas operating over a frequency flat fading channel. To simplify the discussion in this section, only a simple concatenated system is considered, as shown in Fig. 1. The sequel will consider more complex systems with interleavers, but the notation will not change significantly. In the transmitter block diagram shown in Fig. 1, the original information $\{I(k)\}$ goes through the space-time encoder to get L_t streams of data. Each stream of data is used as the input to a CPM modulator. The modulated signals are transmitted simultaneously through L_t transmit antennas. Consequently, the signal received at each receive antenna is a noisy superposition of the L_t transmitted signals corrupted by Rayleigh fading and an independent zero-mean complex additive white Gaussian noise (AWGN) with single-sided power spectral density N_0.

If we limit the CPM scheme to be the same over each transmit antenna and of single modulation index, the signal received by antenna j, $Y_j(t,\mathbf{D})$[1] can be given by [8]

$$Y_j(t,\mathbf{D}) = \sum_{m=1}^{L_t} C_{m,j}(t) X_m(t,\vec{D}_m) + N_j(t) \quad j = 1,\ldots,L_r \tag{1}$$

$$X_m(t,\vec{D}_m) = \sqrt{\frac{E_s}{T}} \exp\left(j\phi(t,\vec{D}_m)\right) \quad m = 1,\ldots,L_t \tag{2}$$

[1]The notation in this paper will be such that random quantities are shown as uppercase characters, whereas deterministic quantities or realizations of random quantities are shown as lower case.

$$\phi(t,\vec{D}_m) = 2\pi h \sum_{l=0}^{k} D_m(l) q(t-lT) \quad kT \le t < (k+1)T \tag{3}$$

where

- $C_{m,j}(t)$ path gain between the mth transmit antenna and the jth receive antenna;
- $D_m(l)$ lth modulation symbol on the mth transmit antenna;
- h modulation index;
- $q(t)$ phase smoothing function;
- T symbol time;
- E_s symbol energy.

Similarly, as in linear modulation, we define the codeword matrix $\mathbf{D}$ in the following:

$$\mathbf{D} = \begin{bmatrix} D_1(1) & \cdots & D_{L_t}(1) \\ D_1(2) & \cdots & D_{L_t}(2) \\ \vdots & \vdots & \vdots \\ D_1(N_c) & \cdots & D_{L_t}(N_c) \end{bmatrix} = \left[\vec{D}_1,\ldots,\vec{D}_{L_t}\right] = \begin{bmatrix} \mathbf{D}(1) \\ \vdots \\ \mathbf{D}(N_c) \end{bmatrix} \tag{4}$$

where N_c is the frame size in terms of symbol time.

CPM is a nonlinear modulation that is defined by a phase trellis. With rational modulation indices, say $h = (2n/p)$, where n and p are relatively prime integers, $\phi(t,\vec{D}_m)$ at each symbol time is constrained to lie on a trellis

$$\phi(t,\vec{D}_m) = \Theta_m(k) + 2\pi h \sum_{l=k-L_q+1}^{k} D_m(l) q(t-lT) \quad kT \le t < (k+1)T \tag{5}$$

where $\Theta_m(k) (\mathrm{mod} 2\pi) \in \Omega_\Theta$, and

$$\Omega_\Theta = \left\{0, \frac{2\pi}{p}, \frac{4\pi}{p}, \ldots, \frac{2(p-1)\pi}{p}\right\}. \tag{6}$$

The state at time kT for each transmit antenna is defined as

$$S_m(k) = [\Theta_m(k), D_m(k-L_q+1),\ldots,D_m(k-1)]. \tag{7}$$

$S_m(k) \in \Omega_{S_m}$ and $||\Omega_{S_m}|| = pM^{L_q-1}$, where $||\Omega_{S_m}||$ denotes the cardinality of Ω_{S_m}. The overall transmitter can therefore be represented by a super trellis with the state at time kT being

$$S(k) = [S_1(k), S_2(k),\ldots,S_{L_t}(k), I(k-\nu),\ldots,I(k-1)] \tag{8}$$

where ν is the memory length of the space-time encoder, and $S(k) \in \Omega_S$ and $||\Omega_S|| = p^{L_t} M^{L_t(L_q-1)+\nu}$. $S(k)$ can be further simplified as

$$S(k) = [\Theta_1(k),\ldots,\Theta_{L_t}(k) \; I(k-L_q-\nu+1),\ldots,I(k-1)]. \tag{9}$$

$||\Omega_S||$ is now reduced to $p^{L_t} M^{L_q+\nu-1}$. Note that delay diversity gives the minimal state number since there is only one phase

state needs to be included in $S(k)$ [9]. In the sequel, delay diversity CPM will be chosen as the inner part of the interleaved system due to its low complexity. We will also denote $\phi(t, \vec{D}_m) = \phi(t, S(k), I(k))$ to indicate the data dependence on the state $S(k)$.

This paper considers the case where the channel gains are spatially independent and time varying. The fading processes are assumed to be independent of each other, with the autocorrelation function being

$$\begin{aligned} R_C(\tau, m1, j1, m2, j2) =& E\left[C_{m1,j1}(t) C^*_{m2,j2}(t-\tau)\right] \\ =& J_0(2\pi f_D \tau)\delta_{m1-m2}\delta_{j1-j2} \end{aligned} \quad (10)$$

where f_D is the Doppler spread, and $m_i = 1, L_t$ and $j_i = 1, L_r$. Since all the spatial channels are modeled with statistically identical and independent models, we denote the single channel correlation with $R_C(\tau)$. Spatial dependence can be incorporated in the model and the corresponding algorithm can be derived (e.g., [20], [21]), but that is not considered in this paper. Additionally, it is assumed that the fading is slow enough so that it remains roughly constant during one symbol period, i.e., $C_{m,j}(t) = C_{m,j}(k)$, $kT \le t < (k+1)T$.

III. K_d-Lag Soft-Output Demodulator

The following notations are useful in deriving the algorithm. $Y_j(k) = \{Y_j(t): kT \le t < (k+1)T\}$ for any time series $\{f(0), f(1), \ldots, f(k)\}$, $\vec{f}(k) = [f(k) \ \ldots \ f(0)]^T$, and boldface denotes signals over all receive antennas, e.g., $\mathbf{c}_j(k) = [c_{1j}(k), \ldots, c_{L_t j}(k)]^T$, $\mathbf{c}(k) = [\mathbf{c}_1^T(k), \ldots, \mathbf{c}_{L_r}^T(k)]^T$, and $\mathbf{Y}(k) = [Y_1(k), \ldots, Y_{L_r}(k)]^T$.

The optimum K_d-lag soft-output receiver produces the metric $p(i(k-K_d)|\vec{\mathbf{y}}(k))$. This probability mass function (PMF) can be written as

$$p\left(i(k-K_d)|\vec{\mathbf{y}}(k)\right) = \sum_{\{\vec{i}(k)\} \in \Gamma(i(k-K_d))} p\left(\vec{i}(k)|\vec{\mathbf{y}}(k)\right) \quad (11)$$

where $\Gamma(i(k-K_d)) = \{\vec{I}(k): I(k-K_d) = i(k-K_d)\}$. The set $\Gamma(i(k-K_d))$ represents all possible trellis trajectories such that $I(k-K_d) = i(k-K_d)$.

The PMF $p(\vec{i}(k)|\vec{\mathbf{y}}(k))$ can be computed in an efficient recursive fashion as in (12), shown at the bottom of the page. Similarly, as in single transmit antenna case, $p(\vec{i}(k-1)|\vec{\mathbf{y}}(k-1))$ will be denoted the sufficient statistic for this recursive algorithm, and $f(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1), \vec{i}(k))$ will be denoted the innovation probability density function (PDF) as it serves very much the same function as the innovation process in recursive filtering. $p(i(k))$ is the *a priori* PMF of the transmitted data symbols.

The most important step is the computing of the innovation density, which will be derived in the Appendix. The expression is given as

$$\begin{aligned} f(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1), \vec{i}(k)) =& C_f \prod_{j=1}^{L_r} \frac{1}{\det(\Sigma_e)\det(\Psi(k))} \\ & \times \exp\left\{\mathbf{x}_j^H(k)\Psi^{-1}\mathbf{x}_j(k) - \hat{\mathbf{c}}_j^H(k)\Sigma_e^{-1}\hat{\mathbf{c}}_j(k)\right\} \end{aligned} \quad (13)$$

with

$$\Psi(k) = \frac{E_s}{N_0}A(k) + \Sigma_e^{-1} \quad (14)$$

$$\mathbf{x}_j(k) = \Sigma_e^{-1}\hat{\mathbf{c}}_j(k) + \frac{E_s}{N_0}\mathbf{q}_j(k) \quad (15)$$

where $\hat{\mathbf{c}}_j(k)$ and Σ_e are the estimated mean and the estimation error covariance matrix of $\mathbf{c}_j(k)$ conditioned on the past information, $\mathbf{q}_j(k)$ are the outputs of the matched filter bank at the jth receive antenna for L_t transmitted signals, and $A(k)$ results from the signal interaction from each transmit antenna.

Approximations must be made to the above algorithm to make it practically feasible. The exact implementation of the above algorithm is impractical since its computation complexity grows exponentially with the frame length due to the infinite memory in the fading processes. In our simulation, we fix the Wiener filter to be of finite length N_d. If N_d is greater than K_d, then decision feedback is used in the Wiener prediction filter for observations for which decisions have been made.

IV. Pilot Design

Pilot design in space-time-coded CPM is a more important issue than that in single transmit antenna systems or linearly modulated multi-antenna systems. Even in single-antenna systems, CPM pilot symbol designs require more thought than in linear modulation due to the modulation memory [26]. In [19], pilot symbol-assisted CPM demodulation using all the observations (pilots plus data) in a single antenna channel is investigated, and the performance gain is not significant compared with that without pilots (blind estimation). In space-time coding with joint demodulation, the initial channel estimates have a significant effect on the frame error rate (e.g., [16]). Consequently, the design of the pilot symbol-assisted CPM to arrive at good initial channel estimates is important.

For the above-mentioned reasons, we propose a frame structure with a preamble of pilot symbols to make the initial channel estimates and a payload with no pilot symbols. This frame structure is shown in Fig. 2. An important quantity for this discussion is the estimation error covariance matrix conditioned on the observed noisy pilot symbols, which we denote Σ_p. The pilot matrix design with linear modulation is investigated in [3] and

$$p\left(\vec{i}(k)|\vec{\mathbf{y}}(k)\right) = \frac{f(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1), \vec{i}(k)) p(\vec{i}(k-1)|\vec{\mathbf{y}}(k-1)) p(i(k))}{\sum_{\vec{i}(k)} f(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1), \vec{i}(k)) p(\vec{i}(k-1)|\vec{\mathbf{y}}(k-1)) p(i(k))}. \quad (12)$$

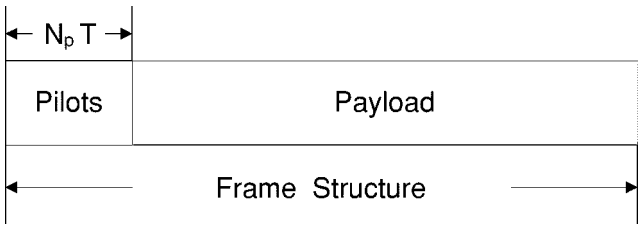

Fig. 2. Frame structure for the space-time CPM system.

[6], where a good design strategy is to minimize the trace of Σ_p while keeping its off-diagonal elements to be zero. It is shown that orthogonal pilot signal designs would provide the desired performance characteristics in quasistatic fading. The same result is also used in [10], and [23]. Due to the phase continuity of CPM signals, a pilot matrix that makes Σ_p diagonal might not be available; therefore, the design is more general than the linear modulation case.

Since Σ_p cannot be forced to be a diagonal matrix, all components of the Σ_p matrix are important to consider. The diagonal components stand for the mean square estimation error of each path gain. The off-diagonal components represent the cross correlation between any two path gain estimation errors. It would seem imprudent to drive the mean square estimation error smaller at the cost of a high correlation in the path gain estimation errors. Consequently, the trace minimization proposed in [3] and [6] is not sufficient as it does not address the correlation between the errors in estimating the path gains. It seems obvious and plausible that a desirable condition for a pilot design would be to minimize the path gain estimation errors and not to have significant correlation between the different path gain estimation errors. Therefore, we consider the squared Frobenius norm $||\Sigma_p||_F^2$ as the cost function to be minimized.

For short preambles, closed-form analytical results can be computed for the estimation error covariance matrix. As with CPM, the interleaver cannot be used before sending the signal to channel; otherwise, phase continuity will be disrupted, and the fading is usually changing slowly in practical scenario; hence, for short pilot preambles, it is reasonable to assume that the fading processes remain constant during the pilot length N_p, and the estimation error covariance matrix can be simplified into

$$\Sigma_p = I - R^H \left(RR^H + \frac{N_0}{E_s} I \right)^{-1} R. \tag{16}$$

Here, with a slight abuse in notation, R is now defined as

$$R = \begin{bmatrix} \vec{\rho}_{11}(N_p-1) & \dots & \vec{\rho}_{L_t 1}(N_p-1) \\ \vdots & \vdots & \vdots \\ \vec{\rho}_{1L_t}(N_p-1) & \dots & \vec{\rho}_{L_t L_t}(N_p-1) \end{bmatrix} \tag{17}$$

where $\vec{\rho}_{ij}(N_p-1)$ is a vector of the time cross correlations between transmitted signal from antenna i and the transmitted signal from antenna j. The computation of each component $\rho_{ij}(k)$ follows the form given in (A.5) in the Appendix. The elements of the matrix R represent the cross correlation of the CPM pilot waveforms transmitted from each of the antennas over the whole preamble time. Note that the size of R is $L_t N_p$ by L_t, and its maximum rank is L_t.

Direct computation of Σ_p is not an easy task and is not necessary as we are only concerned with $||\Sigma_p||_F^2$. Instead, we perform the singular value decomposition (SVD) of R as $R = U\Lambda V^H$, where U is a unitary matrix of size $L_t N_p$ by $L_t N_p$ and V is a unitary matrix of size L_t by L_t. Λ is given in the following:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_{L_t} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}_{L_t N_p \times L_t}. \tag{18}$$

Then, we have

$$||\Sigma_p||_F^2 = ||V^H \Sigma_p V||_F^2 \tag{19}$$

$$= \left\| I - \Lambda^H \left(\Lambda\Lambda^H + \frac{N_0}{E_s} I \right)^{-1} \Lambda \right\|_F^2 \tag{20}$$

$$= \sum_{m=1}^{L_t} \left(\frac{\frac{N_0}{E_s}}{|\lambda_m|^2 + \frac{N_0}{E_s}} \right)^2. \tag{21}$$

From $||\Sigma_p||_F^2$, the following properties can be identified.

1) When R is full rank, $||\Sigma_p||_F^2$ goes to 0 when SNR goes to infinity. If R is not full rank, then $||\Sigma_p||_F^2$ stabilizes at 1, no matter what the SNR is. Consequently, full rank is important for pilot/preamble design. With linear modulation in order for Σ_p to achieve full rank, at least L_t symbols are needed. This is the same result as that given in [22].

2) When SNR is low, the performance loss due to the rank deficiency of R will not be as significant as when SNR is high.

3) Changing the initial phase offset of the transmitted signal will not result in a change in performance since the value of $|\lambda_m|$ will not be changed.

4) The orthogonal pilot design in linear modulation is certainly a special case where Σ_p can be chosen to be diagonal. In [3] and [6] Guey *et al.* show that the normalized Hadamard matrix is a desired pilot matrix and that the form of this matrix is only a function of number of transmit antennas. Usually, the diagonal Σ_p does not exist with CPM since it is hard to make the waveform orthogonal while maintaining the phase continuity. Moreover, the optimal pilot matrix depends on the CPM scheme *and* the number of transmit antennas. Let us take MSK signals with $L_t = 2$ and $N_p = 8$ as an example. One set of the best pilots that gives the minimal $||\Sigma_p||_F^2$ will be

$$\mathbf{D_p} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T. \tag{22}$$

The preamble flips the phase to be the antipodal from the two antennas during the first symbol period, and then maintains this antipodal characteristic through the first half of the preamble. At the halfway point of the preamble, the phase is returned back to being the same value on both antennas and kept that way until the end of the preamble. Actually, this pilot construction works for all full-response CPM schemes with $h = 0.5$ and $L_t = 2$. Note the similarity of this design to the Hadamard matrix design proposed in [6].

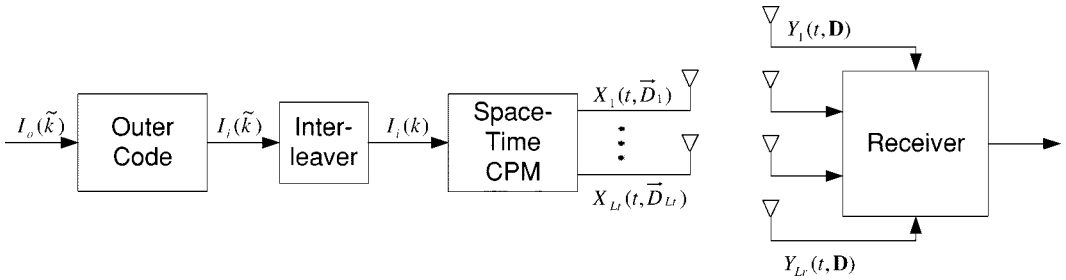

Fig. 3. Block diagram of concatenated space-time CPM.

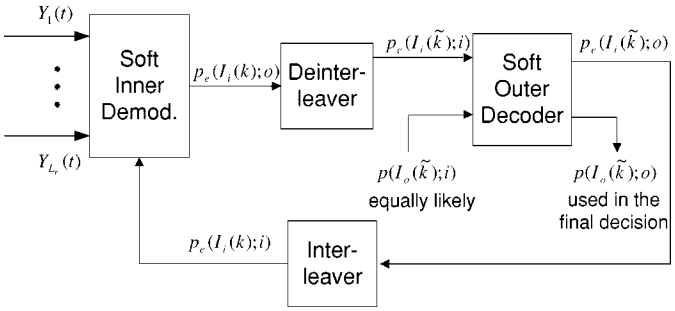

Fig. 4. Block diagram of concatenated space-time CPM demodulator.

V. Interleaved System

An interleaved system can yield improved performance by using iterative detection. The adaptive soft output demodulator derived above makes the "turbo" processing possible. For the uninterleaved system, if we want to achieve high performance, we must resort to complex space-time codes. This is impractical to be implemented due to the computation complexity. It is therefore desirable that the interleaved system be applied where we can concatenate a simple inner space-time CPM with some outer code to get the required frame error rate while keeping the overall complexity moderate. With the recursive nature of CPM signals, space-time CPM is justified as an inner code of a serial concatenated system according to [27]. Moreover, since the Wiener prediction filter is decision-directed $(N_d > K_d)$, the updated soft information after iteration tends to yield more accurate channel estimates. The block diagram of the concatenated space-time CPM is shown in Fig. 3.

Fig. 4 shows the block diagram of the concatenated demodulator. The detection procedure is similar to that for standard serial concatenated codes [28]. Based on the received signals from each receive antenna, $Y_1(t), Y_2(t), \ldots, Y_{L_r}(t)$ and the *a priori* information $p_e(I_i(k);i)$, the inner adaptive demodulator calculates the soft information $p_e(I_i(k);o)$ and passes it to the outer soft output decoder to get the updated *a priori* information, which is used for next iteration. After several exchanges of soft information, the outer decoder calculates $p(I_o(\tilde{k});o)$ and makes the final decision of the information bits.

VI. Simulation Results

In this section, we present some simulation results to verify our soft output algorithm and our pilot design. The algorithm will be parameterized with (K_d, N_d, N_p), where we recall that

K_d decision lag;
N_d Wiener prediction filter length;
N_p preamble length.

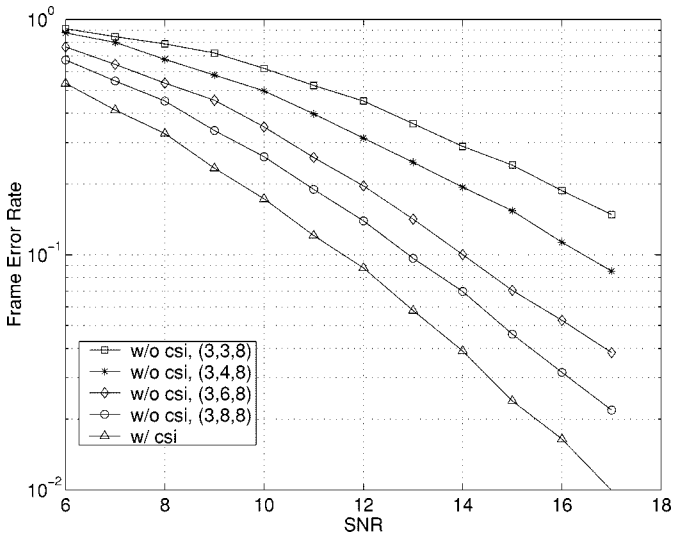

Fig. 5. Simulated FER of MSK with delay diversity. $K_d = 3$, $L_t = 2$, $L_r = 1$, and $f_dT = 0$.

The frame length is chosen to be 130. N_p should be long enough to give good initial estimates. We also set N_p sufficiently larger than $(N_d - K_d)$ to push the decision-directed Wiener prediction filter moving along the correct path at the very beginning. All simulation systems assume binary input, full rate, and normalized signal-to-noise ratio (SNR) versus all transmit/receive antennas; hence, all the performance curves are compared under the same condition. The SNR loss due to pilot symbols is not shown in these figures.

A. Performance Variation With Different K_d and N_d

Different K_d and N_d will result in different performance and different computation complexity. Here, we take MSK with delay diversity in quasistatic fading as an example. The frame error rates of $K_d = 3$ and $K_d = 4$ are shown in Figs. 5 and 6, respectively. We see that with larger K_d, better performance can be obtained, but the complexity of the algorithm also grows exponentially. Increasing N_d improves the performance up to a point. A rule of thumb we observed is that good performance is obtained with $2K_d \leq N_d \leq 4K_d$. If N_d is too large, some degradation can occur due to the data feedback in Wiener prediction filter. Larger N_d will also produce a more complex demodulator since there is an L_tN_d by L_tN_d matrix inversion in the algorithm. Note that unlike in the single transmit antenna case, where the Wiener prediction coefficients and the estimation error covariance matrix can be precomputed offline [19], [20], in space-time coding, we need to compute

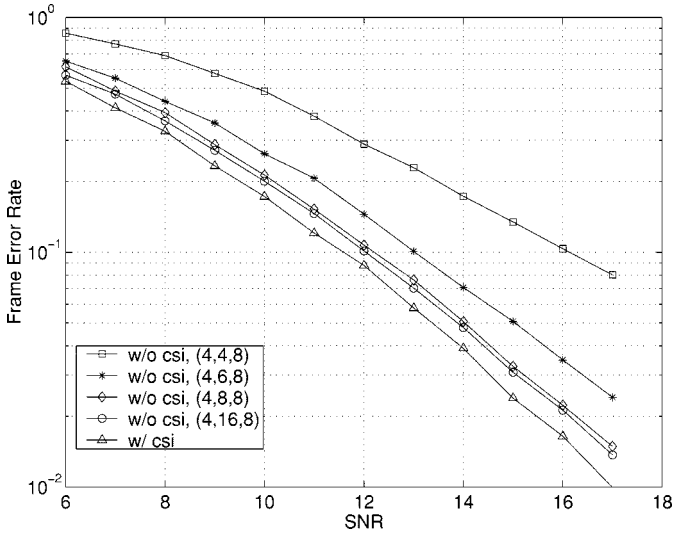

Fig. 6. Simulated FER of MSK with delay diversity. $K_d = 4$, $L_t = 2$, $L_r = 1$, and $f_dT = 0$.

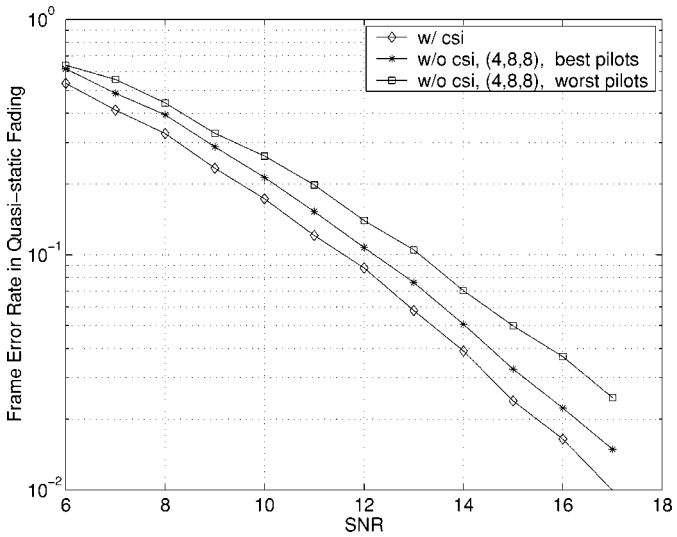

Fig. 7. Simulated FER of MSK with delay diversity using different pilot sequence. $L_t = 2$, $L_r = 1$, and $f_dT = 0$.

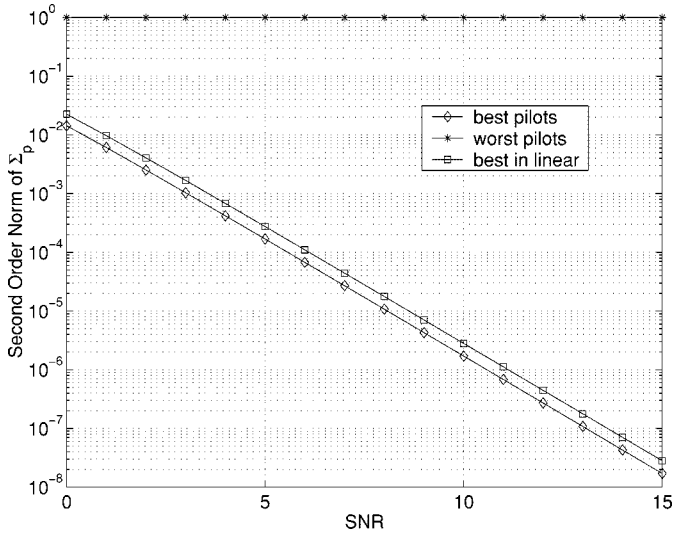

Fig. 8. Cost function with different pilot sequence. $L_t = 2$, $L_r = 1$, and $f_dT = 0$.

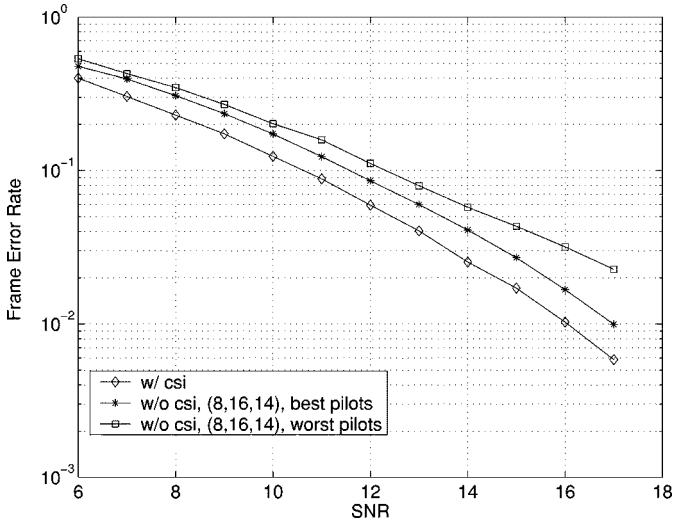

Fig. 9. Simulated FER of binary 1RC with $g(5,7)$. $L_t = 2$, $L_r = 1$, and $f_dT = 0$.

these values online. This phenomenon can be explained by examining (A.11) and (A.12). They are the functions of signal correlation from each transmit antenna, which, in turn, depends on the particular data path. An alternative is to compute $\left(RC_cR^H + N_0/E_sI\right)^{-1}$ for each data path beforehand and store these values in a big matrix.

B. Performance Comparison With Different Pilot Matrix

The frame error rate is also affected by the choice of pilot symbol preamble sequences. Fig. 7 gives the performance curve of MSK with delay diversity in quasistatic fading with different pilots. The best pilot matrix is as in (22), whereas the worst pilot matrix corresponds to the all zero pilots on both antennas. From this figure, we can see that by using the designed pilots, we can achieve near-coherent performance. On the other hand, if the pilot words are not chosen properly, performance degradation will occur, especially at high SNR, which is consistent with our previous analysis in Section IV. The different performance resulting from different initial channel estimates with linear modulation is also observed in [16].

The cost functions with different pilot matrices are plotted in Fig. 8. The lowest cost results from the best pilots, and a cost of around 1 results from the worst pilots. It is interesting to note that the channel estimation from the preamble in the worst case appears to have an error floor, and yet, the frame error rate does not exhibit an observable error floor. This apparent contradiction is explained by both the demodulator's ability to adaptively recover from bad initial estimates and the fact that most often in a fading environment, the instantaneous SNR is very large. Hence, only a fraction of the frames that start with bad channel estimates will have frame errors. Consequently, the frame error rate has no observable error floor.

The frame error rate of binary 1RC with $g(5,7)$ as the space-time encoder is shown in Fig. 9. $g(5,7)$ denotes the generator polynomial of the encoder expressed in octal form,

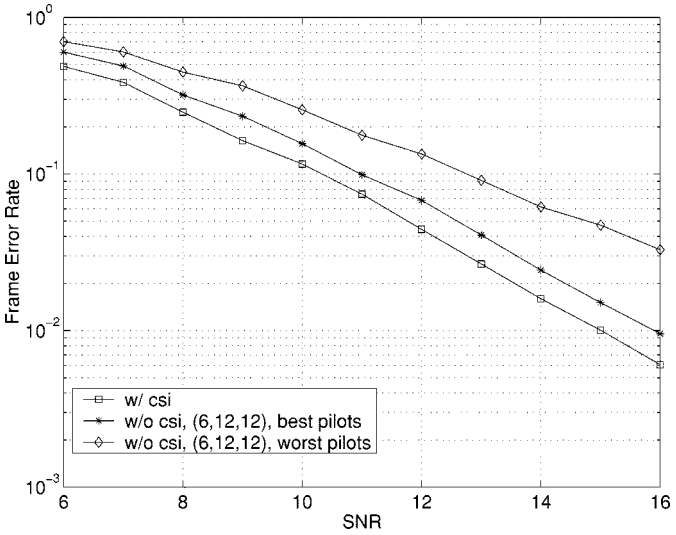

Fig. 10. Simulated FER of MSK with delay diversity using different pilot sequence. $L_t = 3$, $L_r = 1$, and $f_dT = 0$.

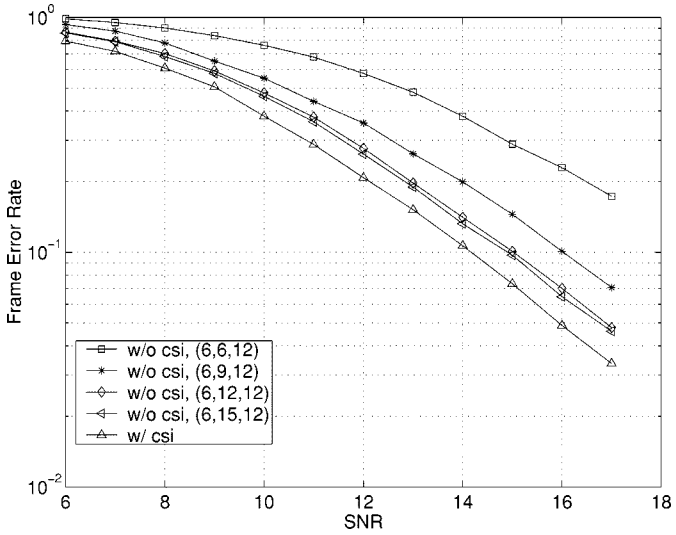

Fig. 11. Simulated FER of MSK with delay diversity. $K_d = 6$, $L_t = 2$, $L_r = 1$, and $f_dT = 0.008$.

and its guaranteed full diversity has been proved in [29]. Once again, we can see with the properly designed pilots, the performance is close to that given perfect channel state information, whereas the improper pilots will lead to performance loss.

The effect of initial channel estimates is even more significant as the number of transmit antennas gets larger. Fig. 10 shows the frame error rate of MSK with delay diversity and $L_t = 3$. Matrix R in (17) with the worst-case pilot matrix has rank 1, whereas its rank with the best pilot matrix is 3. Consequently, the performance deterioration of the worst pilots is much greater than the two transmit antenna case.

C. Performance in Time-Varying Fading

The performance of the algorithm in time-varying fading is demonstrated in Figs. 11–13. These three figures show the frame error rate of MSK with delay diversity. The normalized Doppler spread is chosen to be $f_DT = 0.008$. In this case, we can still observe the similar performance variation with different K_d and N_d, as in quasistatic fading. The performance gap between the worst pilots and the best pilots tightens a little bit compared with that in quasistatic fading since the pilot symbols are designed under the approximation that the fading stays constant over pilots, but a similar trend can still be observed. As we have pointed out earlier, the EM algorithm does not work well in fast fading due to its iterative nature. The herein proposed demodulation algorithm estimates the channel at the very beginning through the pilot symbols; it then keeps tracking the channel over time, and thus, it can perform well, even with $f_DT = 0.008$ and higher. This performance would also be typical with the PSP-based algorithms in [15] and [16], but iterative demodulation would not be feasible due to the hard decisions in these algorithms.

D. Performance in Interleaved System

The performance of the interleaved system is investigated by taking an example of delay diversity 4-ary 1RC with the outer binary code being $g(5, 7)$. Note that the overall system still has full rate. Delay diversity gives the minimal complexity for inner space-time CPM, whereas $g(5, 7)$ is the 1/2 binary convolu-

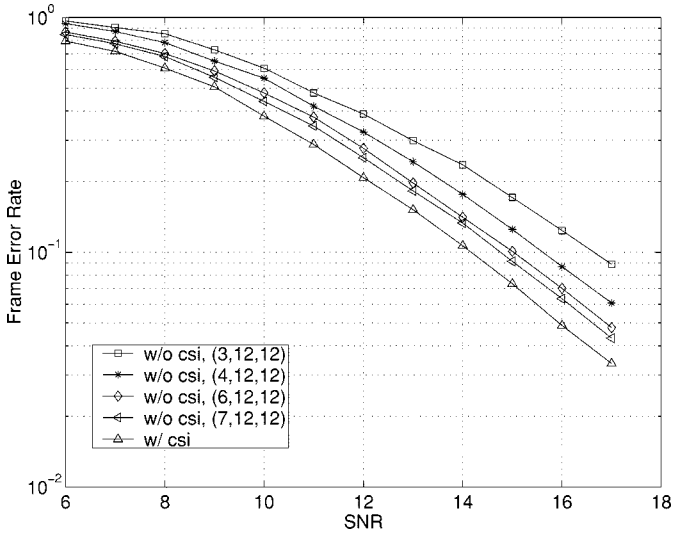

Fig. 12. Simulated FER of MSK with delay diversity. $N_d = 12$, $L_t = 2$, $L_r = 1$, and $f_dT = 0.008$.

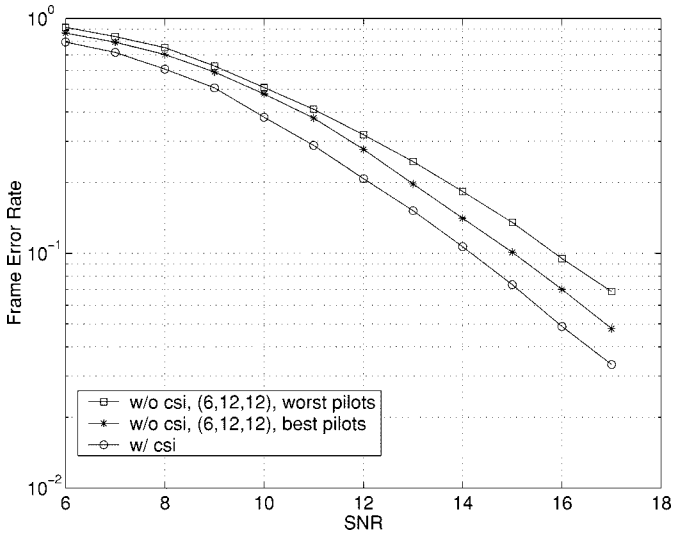

Fig. 13. Simulated FER of MSK with delay diversity using different pilot sequence. $L_t = 2$, $L_r = 1$, and $f_dT = 0.008$.

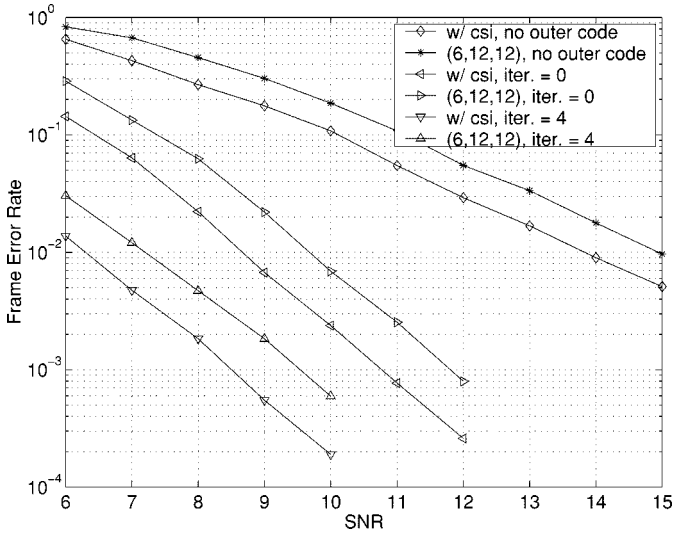

Fig. 14. Simulated FER of Delay Diversity 1RC with $g(5,7)$ in serial concatenated system. $L_t = 2$, $L_r = 2$, and $f_D T = 0.008$.

tional codes having the optimal d_{free} with memory length 2. According to [27], a good serial concatenated system should choose an outer code with a large minimum Hamming distance. From Fig. 14, we can see that performance is greatly improved with the interleaved structure since time diversity is exploited, and the diversity gain is significant with multiple antennas even when the frame length is small, which is attractive for delay-sensitive applications. Even with as few as four iterations of soft information exchange between the inner space-time demodulator and the outer decoder, a larger coding gain can be obtained. The reason why we chose number of iterations to be 4 is that in the fading channel, most improvement is achieved by the first iteration, and four iterations give a good compromise between performance and complexity. This figure also indicates that the adaptive receiver achieves decent performance in interleaved system, and extension of the algorithm designed for one receive antenna to multiple receive antennas is fairly straightforward.

VII. Conclusion

In this paper, we proposed a soft output receiver structure for joint channel estimation and data detection in multiple antenna systems based on trellis structure of the space-time CPM system and Wiener prediction filter. Pilot symbols are inserted at the very beginning to give good initial estimates of the channel. The design of preamble pilot symbols that achieve optimum performance is investigated. This soft receiver is further applied in a serial concatenated space-time CPM system to achieve better performance through iterative decoding. Simulation results show that this proposed algorithm works well in quasistatic fading as well as in time-varying fading.

Appendix

In this Appendix, we calculate the innovation PDF $f(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k))$ in detail. Using Bayes rule and the trellis structure of the ST-CPM system, the innovation PDF can be written as

$$f\left(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k)\right)=\int f\left(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k),\mathbf{c}(k)\right)\times f\left(\mathbf{c}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k)\right)d\mathbf{c}(k). \quad \text{(A.1)}$$

In order to write out the first term in the above integration (A.1), a matched filter needs to be computed for each possible path $\vec{i}(k)$ through the phase trellis, and it is given as

$$\begin{aligned} q_{m,j}(k;\vec{i}(k)) =& \frac{1}{\sqrt{E_s T}} \times \int_{kT}^{(k+1)T} y_j(t) e^{-j\phi\left(t,\vec{d}_{\vec{i}(k),m}\right)} dt \quad \text{(A.2)} \\ =& \frac{1}{T}\int_{kT}^{(k+1)T}\sum_{l=1}^{L_t} c_{l,j}(t) e^{j\phi(t,\vec{d}_l)} \times e^{-j\phi\left(t,\vec{d}_{\vec{i}(k),m}\right)} dt + n_{m,j}(k) \\ =& \sum_{l=1}^{L_t} c_{l,j}(k)\rho_{l,m}(\mathbf{d},\mathbf{d}_{\vec{i}(k)},k) + n_{m,j}(k) \quad \text{(A.3)} \end{aligned}$$

where

$$n_{m,j}(k)=\int_{kT}^{(k+1)T} n_{m,j}(t) e^{-j\phi\left(t,\vec{d}_{\vec{i}(k),m}\right)} dt. \quad \text{(A.4)}$$

$\vec{d}_{\vec{i}(k),m}$ stands for the stream sent on the mth transmit antenna given a particular data path $\vec{i}(k)$, and $\rho_{l,m}(\mathbf{d}_\alpha,\mathbf{d}_\beta,k)$ is the cross correlation function of the various CPM waveforms, i.e.,

$$\rho_{l,m}(\mathbf{d}_\alpha,\mathbf{d}_\beta,k)=\frac{1}{T}\int_{kT}^{(k+1)T} e^{j\phi\left(t,\vec{d}_{\alpha,l}\right)} e^{-j\phi\left(t,\vec{d}_{\beta,m}\right)} dt. \quad \text{(A.5)}$$

When $\mathbf{d}_\alpha = \mathbf{d}_\beta$, we simplify $\rho_{l,m}(\mathbf{d}_\alpha,\mathbf{d}_\beta,k)$ as $\rho_{l,m}(\mathbf{d}_\alpha,k)$.

Hence, the first term can be written as in [30]

$$\begin{aligned} & f\left(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k),\mathbf{c}(k)\right) \\ &= C_f \prod_{j=1}^{L_r} \exp\left\{2\frac{E_s}{N_0}\Re\left[\mathbf{q}_j^H(k)\mathbf{c}_j(k)\right] - \frac{E_s}{N_0}\mathbf{c}_j^H(k)A(k)\mathbf{c}_j(k)\right\} \quad \text{(A.6)} \end{aligned}$$

where[2]

$$\mathbf{q}_j(k)=[q_{1,j}(k),\ldots,q_{L_t,j}(k)]^T \quad \text{(A.7)}$$

$$\mathbf{c}_j(k)=[c_{1,j}(k),\ldots,c_{L_t,j}(k)]^T \quad \text{(A.8)}$$

$$A(k)=\begin{bmatrix} \rho_{11}(\mathbf{d}(\vec{i}(k)),k) & \cdots & \rho_{L_t 1}(\mathbf{d}(\vec{i}(k)),k) \\ \vdots & \vdots & \vdots \\ \rho_{1L_t}(\mathbf{d}(\vec{i}(k)),k) & \cdots & \rho_{L_t L_t}(\mathbf{d}(\vec{i}(k)),k) \end{bmatrix}. \quad \text{(A.9)}$$

[2]Note, for simplicity, that we dropped the dependence of $q_{i,j}(k)$ on information sequence $\vec{i}(k)$ in the notation, i.e., $q_{i,j}(k;\vec{i}(k))$ is simplified as $q_{i,j}(k)$

The second term in (A.1) involves estimating the channel state information through a Wiener prediction filter. Due to the joint Gaussian nature of the channel state information and the past matched filter outputs, the CSI conditioned on past information has the following probability distribution:

$$f\left(\mathbf{c}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k)\right)=\prod_{j=1}^{L_r}\frac{1}{\pi^{L_t}\det(\Sigma_e)}\times\exp\left\{-[\mathbf{c}_j(k)-\hat{\mathbf{c}}_j(k)]^H\Sigma_e^{-1}[\mathbf{c}_j(k)-\hat{\mathbf{c}}_j(k)]\right\} \quad \text{(A.10)}$$

with $\hat{\mathbf{c}}_j(k)=E[\mathbf{C}_j(k)|\vec{\mathbf{q}}_j(k-1)]$ being the estimated mean and $\Sigma_e=E[(\mathbf{C}_j(k)-\hat{\mathbf{c}}_j(k))(\mathbf{C}_j(k)-\hat{\mathbf{c}}_j(k))^H]$ being the estimation error covariance matrix. More specifically

$$\hat{\mathbf{c}}_j(k)=B^HR^H\left(RC_cR^H+\frac{N_0}{E_s}I\right)^{-1}\times\vec{\mathbf{q}}_j(k-1) \quad \text{(A.11)}$$

$$\Sigma_e=I-B^HR^H\left(RC_cR^H+\frac{N_0}{E_s}I\right)^{-1}RB \quad \text{(A.12)}$$

where R stands for the signal correlation from each transmit antenna, which is given by

$$R=[R_{ij}] \quad \text{(A.13)}$$

$$R_{ij}=\text{diag}\left\{\rho_{ij}(\mathbf{d}(\vec{i}(k)),k-1),\ldots,\rho_{ij}(\mathbf{d}(\vec{i}(k)),0)\right\}^H. \quad \text{(A.14)}$$

B and C_c result from the channel autocorrelation since the fading process has memory. B is given as

$$B=\begin{bmatrix}\vec{v} & \ldots & \vec{0}\\ \vdots & \vdots & \vdots\\ \vec{0} & \ldots & \vec{v}\end{bmatrix} \quad \text{(A.15)}$$

$$\vec{v}=[\,J_0(2\pi f_D\cdot 1) \quad \ldots \quad J_0(2\pi f_Dk)\,]^T \quad \text{(A.16)}$$

and

$$C_c=\begin{bmatrix}C_d & \ldots & 0\\ \vdots & \vdots & \vdots\\ 0 & \ldots & C_d\end{bmatrix} \quad \text{(A.17)}$$

$$C_d=\begin{bmatrix}1 & \ldots & J_0(2\pi f_D\cdot(k-1))\\ J_0(2\pi f_D\cdot(-1)) & \ldots & J_0(2\pi f_D\cdot(k-2))\\ \vdots & \vdots & \vdots\\ J_0(2\pi f_D\cdot(-k+1)) & \ldots & 1\end{bmatrix}. \quad \text{(A.18)}$$

Therefore, we have the innovation PDF

$$f(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k))=C_f\prod_{j=1}^{L_r}\int\exp\left\{2\frac{E_s}{N_0}\Re\left[\mathbf{q}_j^H(k)\mathbf{c}_j(k)\right]-\frac{E_s}{N_0}\mathbf{c}_j^H(k)A(k)\mathbf{c}_j(k)\right\}\times\frac{1}{\pi^{L_t}\det(\Sigma_e)}\times\exp\left\{-(\mathbf{c}_j(k)-\hat{\mathbf{c}}_j(k))^H\Sigma_e^{-1}\times(\mathbf{c}_j(k)-\hat{\mathbf{c}}_j(k))\right\}d\mathbf{c}_j(k). \quad \text{(A.19)}$$

Getting the closed form of (A.19) is crucial in making the algorithm feasible. It can be shown that the exponent in the integrand of (A.19) can be written as

$$-\left(\mathbf{c}_j(k)-\Psi^{-1}(k)\mathbf{x}_j(k)\right)^H\Psi(k)\left(\mathbf{c}_j(k)-\Psi^{-1}(k)\mathbf{x}_j(k)\right)+\mathbf{x}_j^H(k)\Psi^{-1}(k)\mathbf{x}_j(k)-\hat{\mathbf{c}}_j^H(k)\Sigma_e^{-1}\hat{\mathbf{c}}_j(k) \quad \text{(A.20)}$$

where $\Psi(k)$ and $\mathbf{x}_j(k)$ are defined in (14) and (15). The integral in (A.19) is now easily solved by making use of the fact that

$$\int\frac{e^{-\left(\mathbf{c}_j(k)-\Psi^{-1}(k)\mathbf{x}_j(k)\right)^H\Psi(k)\left(\mathbf{c}_j(k)-\Psi^{-1}(k)\mathbf{x}_j(k)\right)}}{\pi^{L_t}\det(\Psi^{-1}(k))}d\mathbf{c}_j(k)$$

is the integral of the PDF of a complex Gaussian vector and, thus, equals one. Consequently, the closed form of (A.19) can be expressed as

$$f\left(\mathbf{y}(k)|\vec{\mathbf{y}}(k-1),\vec{i}(k)\right)=C_f\prod_{j=1}^{L_r}\frac{1}{\det(\Sigma_e)\det(\Psi(k))}\times\exp\left\{\mathbf{x}_j^H(k)\Psi^{-1}\mathbf{x}_j(k)-\hat{\mathbf{c}}_j^H(k)\Sigma_e^{-1}\hat{\mathbf{c}}_j(k)\right\}.$$

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna gaussian channels," AT&T Bell Labs, 1995.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[3] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, Atlanta, GA, 1996, pp. 136–140.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, Autumn 1996.

[6] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. COM-47, pp. 527–537, Apr. 1999.

[7] T. Aulin, N. Rydbeck, and C. W. Sundberg, "Continuous phase modulation—Part I and Part II," *IEEE Trans. Commun.*, vol. COM-29, pp. 196–225, Mar. 1981.
[8] J. B. Anderson, T. Aulin, and C. W. Sundberg, *Digital Phase Modulation*. New York: Plenum, 1986.
[9] X. Zhang and M. P. Fitz, "Space-time coding for Rayleigh fading channels in CPM system," in *Proc. of 38th Annu. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2000.
[10] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility and multiple paths," *IEEE Trans. Commun.*, vol. 47, pp. 199–207, Feb. 1999.
[11] C. Cozzo and B. L. Huges, "Joint channel estimation and data detection in space-time communications," in *Proc. Int. Conf. Commun.*, vol. 1, 2000, pp. 287–291.
[12] S. Perreau and L. B. White, "Channel estimation and symbol detection for space-time codes in mobility conditions," in *Proc. Int. Seminar Broadband Commun.*, Zurich, Switzerland, 2000, pp. 105–108.
[13] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, pp. 47–60, Nov. 1996.
[14] R. Raheli, A. Polydoros, and C.-K. Zhou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb./Mar./Apr. 1995.
[15] Y. Xue and X. Zhu, "PSP decoder for space-time trellis code based on accelerated self-tuning LMS algorithm," *Electron. Lett.*, vol. 36, no. 17, pp. 1472–1474, Aug. 2000.
[16] C. Cozzo and B. L. Hughes, "An adaptive receiver for space-time trellis codes based on per-survivor processing," *IEEE Trans. Commun.*, 2001, submitted for publication.
[17] Z. Liu, X. Ma, and G. Giannakis, "Space-time coding and Kalman filtering for diversity transmission through time-selective fading channels," in *Proc. Milcom*, Dec. 2000, pp. 382–386.
[18] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct 1998.
[19] R. Balasubramanian and M. P. Fitz, "Soft output detection of CPM signals in frequency flat, Rayleigh fading," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1145–1152, July 2000.
[20] X. Zhang and M. P. Fitz, "Soft output diversity combining for CPM signals over space-time correlated Rayleigh fading channels," in *Proc. Int. Conf. Commun.*, 2001.
[21] A. A. Hansson and T. M. Aulin, "Iterative array detection of CPM over continuous-time Rayleigh fading channels," in *Proc. Int. Conf. Commun.*, vol. 7, 2001, pp. 2221–2225.
[22] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, Apr. 2000, submitted for publication.
[23] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Oct. 1998.
[24] K. R. Narayanan and G. L. Stüber, "Performance of trellis coded CPM with iterative demodulation and decoding," in *Proc. Globecom*, vol. 5, 1999, pp. 2346–2351.
[25] P. Moqvist and T. M. Aulin, "Serially concatenated continuous phase modulation with iterative decoding," *IEEE Trans. Commun.*, vol. 49, pp. 1901–1915, Nov. 2001.
[26] P. Ho and J. H. Kim, "On pilot symbol assisted detection of CPM schemes operating in fast fading channels," *IEEE Trans. Commun.*, vol. 44, pp. 337–347, Mar. 1996.
[27] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 909–926, May 1998.
[28] ——, "A Soft-Input Soft-Output Maximum a Posteriori (MAP) Module to Decode Parallel and Serial Concatenated Codes," *TDA Progress Rep.*, vol. 42, Nov. 1996.
[29] X. Zhang and M. P. Fitz, "Space-time code design with CPM transmission," in *Proc. of Int. Symp. Inform. Theory*, 2001.
[30] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

Xiaoxia Zhang received the B.S. and M.S. degrees from the University of Science and Technology of China, Hefei, in 1994 and 1997, respectively, and the Ph.D. degree from The Ohio State University, Columbus, in 2002, all in electrical engineering.

Her research interests include continuous phase modulation, space-time coding, iterative detection, and channel estimation in wireless communications. She is currently working with the Standards Group, Qualcomm, San Diego, CA.

Michael P. Fitz (S'86–M'89) received the B.E.E. degree (summa cum laude) from the University of Dayton, Dayton, OH, in 1983 and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1984 and 1989, respectively.

From 1983 to 1989, he worked as a communication systems engineer for Hughes Aircraft and TRW Inc. and had a fellowship for his graduate education. In 1989, he ventured into academia and has been on the faculty at Purdue University, West Lafayette, IN, and The Ohio State University (OSU), Columbus. He is currently a professor with the Department of Electrical Engineering, University of California, Los Angeles. His research is in the broad area of statistical communication theory and experimentation. His research group is currently interested in the theory of space-time modems and operates an experimental wireless wide area network.

Dr. Fitz received the 2001 IEEE Communications Society Leonard G. Abraham Prize Paper Award in the Field of Communications Systems.