Robust Joint Channel Estimation and Symbol Detection over MIMO Channels Using EM Algorithm

Fu-Hsuan Chiu, Sau-Hsuan Wu and C.-C. Jay Kuo
Department of Electrical Engineering-System, University of Southern California
Los Angeles, California 90089-2564, USA
(E-mail: fchiu@usc.edu, sauhsuaw@usc.edu, cckuo@sipi.usc.edu)

***Abstract*— An expectation maximization (EM) algorithm for joint channel tracking and symbol detection at the receiver in a multi-input multi-output (MIMO) frequency-selective fading environment is proposed in this research. Based on the recursive EM procedure in conjunction with the BCJR algorithm, we develop an iterative algorithm that performs the minimum mean squared error (MMSE) channel estimation and the maximum *a posterior* (MAP) probability symbol detection jointly. The performance of the proposed algorithm is evaluated via simulation and compared with that of Kalman filtering with hard decision feedback. It is also demonstrated that the new algorithm has robust performance in the presence of a severe channel model mismatch.**

I. INTRODUCTION

Space-time coding offers a bandwidth- and power-efficient solution to wireless communications. It exploits multiple transmit and multiple receive antennas to combat fading channels [1]. Although the code design in the original work of Tarokh *et al.* [1] focused primarily on quasi-static flat fading channels, there have been extensions to address more practical models such as time-varying frequency-selective fading channels in recent years [2]. If the channel variation within a transmission block is substantially large, it is important to develop an effective channel tracking scheme to make symbol detection more reliable. This motivates our research on joint channel tracking and symbol detection for such channels.

There have been several schemes proposed for joint maximum likelihood sequence detection (MLSD) and maximum *a posterior* (MAP) or minimum mean squared error (MMSE) channel estimation. Most of these algorithms can be shown to be special applications of the Expectation Maximization (EM) algorithm [3]. Among them, the most well known example is the per-survivor processing (PSP) technique [4]. A drawback of the PSP technique is that its complexity increases exponentially with the channel memory length in time-varying fading channels. Adaptive Bayesian and EM-based detectors for single-in-single-output frequency-selective fading channels were proposed in [5], where the performance of the coded systems was shown to be greatly enhanced by the soft decision-directed channel estimators. A unified structure was proposed in [6] for joint channel tracking and symbol detection in multipath fading channels based on the recursive EM algorithm. Furthermore, in conjunction with the BCJR algorithm, a recursive procedure was presented as a message-passing scheme on a cyclic graph. Here, we extend this result in [6] to multiple-input-multiple-output (MIMO) channels.

The rest of this paper is organized as follow. A general MIMO system over time-varying frequency-selective fading channels is discussed in Sec. II. We separate the space-time channel matrix into two parts according to their temporal characteristics: the time-invariant angle-delay response and the time-varying multipath gain. An efficient way to estimate the angle-delay pattern was given in [7], and it can be obtained in the training phase in the very beginning of transmission. A joint multipath gain tracking and symbol detection scheme using the EM algorithm combined with soft decoding is proposed in Sec. III. The performance of the proposed scheme is evaluated and contrasted with Kalman filter with hard decision feedback in Sec. IV. Concluding remarks are given in Sec. V.

II. CHANNEL MODEL

Consider a MIMO system with L_t-input, L_r-output antennas over time-varying frequency-selective channels. Then, the baseband multipath channel between transmitter k and the receiver of the base station can be modeled as a single-input multiple-output channel with the following vector impulse response:

$$\underline{h}(t) = \sum_{l=1}^{L_p} \mathbf{a}(\phi_{kl}) h_{kl}(t) \delta(t - \tau_{kl}), \tag{1}$$

where L_p is the number of paths in each transmitter's channel, $h_{kl}(t)$ and τ_{kl} are the complex fading gain and the delay of the lth path of the kth transmitter, respectively, and $\mathbf{a}(\phi_{kl})$ is the corresponding array response vector determined by the array geometry and the angle of arrival ϕ_{kl}. At each receive antenna, the baseband received signal is passed through the matched filter and sampled at the symbol rate. The temporal support of the channel is assumed to be $[0, L_c T]$. The discrete-time model can be written as

$$\mathbf{y}_m = \mathbf{H}_m \mathbf{x}_m + \mathbf{n}_m, \tag{2}$$

0-7803-9152-7/05/$20.00

where $\mathbf{y}_m = [y_1(m), \cdots, y_{L_r}(m)]^T$ is the $L_r \times 1$ vector collecting the outputs from all receive antennas at time m; $\mathbf{x}_m = [x_1(m), \cdots, x_{L_t}(m), \cdots, x_1(m-L_c+1), \cdots, x_{L_t}(m-L_c+1)]^T$ is the $L_tL_c \times 1$ transmitted vector; $\mathbf{n}_m$ is the $L_r \times 1$ zero-mean complex colored Gaussian noise vector with covariance $\mathbf{C}$, and the space-time channel matrix $\mathbf{H}_m$ can be decomposed as [7]

$$\begin{aligned}\mathbf{H}_m &= \left[\mathbf{a}(\phi_{11}) \cdots \mathbf{a}(\phi_{L_tL_p})\right] \begin{bmatrix} h_{11}(m) \ \mathbf{0} & \\ \ddots & \\ \mathbf{0} \ h_{L_tL_p}(m) \end{bmatrix} \begin{bmatrix} \mathbf{g}^T(\tau_{11}) \\ \vdots \\ \mathbf{g}^T(\tau_{L_tL_p}) \end{bmatrix} \\ &= \mathbf{A}(\underline{\phi})\mathrm{diag}(\mathbf{h}_m)\mathbf{G}^T(\underline{\tau}),\end{aligned}$$

where $\mathbf{h}_m = [h_{11}(m) \cdots h_{L_tL_p}(m)]^T$ is the multipath CSI vector at time m and $\mathbf{g}(\tau_{ij})$ is the $L_tL_c \times 1$ vector obtained from a zero vector with the $((k-1) \times L_t + i)$-th element replaced by the k-th element of sampled delay waveform vector of the convolution of the transmitted pulse and the matched filter of the jth-path of the i-th transmitter. This model is also applied to diversity arrays by setting spatial correlation matrix $\mathbf{A}(\underline{\phi})$ to an identity matrix. For time-varying channels, the angle of arrival and the time delay of each path are more stationary than the fading channel gain. Thus, we assume that the spatial matrix $\mathbf{A}(\underline{\phi})$ and the temporal matrix $\mathbf{G}(\underline{\tau})$ are time-invariant during M symbols while the channel gain $\mathbf{h}_m$ changes symbol by symbol. We can represent the time variation of the MIMO channel by rearranging the vector form as

$$\begin{aligned}\mathbf{y}_m &= (\mathbf{x}_m^T \otimes \mathbf{I}_{L_r})\mathrm{vec}(\mathbf{H}_m) + \mathbf{n}_m \\ &= (\mathbf{x}_m^T \otimes \mathbf{I}_{L_r})(\mathbf{G}(\underline{\tau}) \odot \mathbf{A}(\underline{\phi}))\mathbf{h}_m + \mathbf{n}_m \\ &= \mathbf{T}_m\mathbf{h}_m + \mathbf{n}_m, \qquad (3)\end{aligned}$$

where $\mathrm{vec}(\cdot)$ is the operator of stacking columns of a matrix into a vector and $\otimes$ and $\odot$ denote the Kronecker and the column-wise Kronecker products, respectively. We assume that the multipath fading gain $h_{kl}(m)$ satisfies the wide-sense-stationary-uncorrelated-scattering (WSSUS) model, which is a zero-mean WSS complex Gaussian process, uncorrelated with any other $h_{k'l'}(m')$. A common way to model the time evolution dynamics of a vector process is to use the autoregressive moving average (ARMA) filter. Let us define $\underline{\mathbf{h}}_m \equiv [\mathbf{h}_m^T, ..., \mathbf{h}_{m-L_h+1}^T]^T$, where L_h is the order of the channel model. The channel dynamics is modeled by a state-space form given by

$$\begin{aligned}\underline{\mathbf{h}}_m &= \underline{\mathbf{F}\mathbf{h}}_{m-1} + \underline{\mathbf{B}}\mathbf{v} \\ &= \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \mathbf{F}_{L_h} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \underline{\mathbf{h}}_{m-1} + \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{v}, \qquad (4)\end{aligned}$$

where $\mathbf{F}_i$, $i = 1, \cdots, L_h$ and $\mathbf{B}$ is of dimension $L_tL_p \times L_tL_p$, and $\mathbf{v}$ is a zero-mean gaussian random vector. Due to the WSSUS assumption, $\mathbf{F}_i$ and $\mathbf{B}$ must be diagonal. The diagonal entries can be selected by exploiting the correlation-matching property and solving Yule-Walker's equations in order to match the physical model. Although the approximation accuracy can be improved by increasing the order L_h, the complexity of the tracking algorithm given in the next section will also increase. On the other hand, we will show in Section IV that the proposed algorithm is robust to severely channel model mismatch. Notice that $\mathbf{A}(\underline{\phi})$ and $\mathbf{G}(\underline{\tau})$ have been introduced to account for the spatial and temporal correlations, respectively, and are assumed known from a training phase [7].

III. Proposed Joint Channel Estimation and Symbol Detection Algorithm

The proposed algorithm is shown in Fig. 1, where the EM algorithm is used to obtain sequential MMSE channel estimators and the soft decoding module is used to provide the required posterior probability as well as symbol detection.

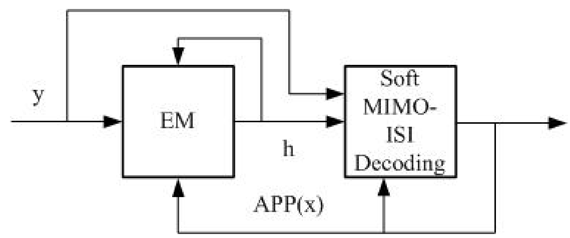

Fig. 1. The block diagram of the proposed joint channel estimator and symbol detection based on EM algorithm and soft decoding.

A. Recursive EM Algorithm

Let $\mathcal{C}_m$, $\mathcal{I}_m$ and θ_m denote the complete data, incomplete data and channel parameters at time m, respectively. Then, the two steps of the EM algorithm can be written as

1) E step: computing the Kullback-Leibler (K-L) measure

$$Q_m(\theta_m|\hat{\theta}_{m|m}^{(l-1)}) = \mathrm{E}\{\log p(\mathcal{C}_m|\theta_m)|\mathcal{I}_m, \hat{\theta}_{m|m}^{(l-1)}\}. \qquad (5)$$

2) M step: maximizing the K-L measure for the new estimate

$$\hat{\theta}_{m|m}^{(l)} = \arg\max_{\theta_m} Q_m(\theta_m|\hat{\theta}_{m|m}^{(l-1)}), \qquad (6)$$

where "$|m$" denotes using information up to time m and l is the total number of iterations at each time step. Let $\mathcal{Y}_m = [\mathbf{y}_m^T, \cdots, \mathbf{y}_0^T]^T$ be the received information, $\mathcal{X}_m = \{\mathbf{x}_m, \cdots, \mathbf{x}_0\}$ be the transmitted data up to time m. Let $\mu_{i|m} = \mathrm{E}\{\mathbf{h}_i|\mathcal{Y}_m, \hat{\theta}_{m|m}^{(l-1)}\}$, $\underline{\mu}_{i|m} = \mathrm{E}\{\underline{\mathbf{h}}_i|\mathcal{Y}_m, \hat{\theta}_{m|m}^{(l-1)}\}$, and $\Sigma_{i,j|m}$, $i, j = 0, \cdots, m$, be the cross-covariance matrix between $\mathbf{h}_i$ and $\mathbf{h}_j$. For the convenience of expression, we define $\hbar_m = [\mathbf{h}_m^T, \cdots, \mathbf{h}_0^T]^T$ with conditional mean $\mathcal{U}_{m|m} = [\mu_{m|m}^T, \cdots, \mu_{0|m}^T]^T$ and conditional covariance matrix $\Gamma_{i|m} = [\Sigma_{i-j,i-l}]_{j+1,l+1}$, which is a block matrix with the block at the $(j+1)$th row and $(l+1)th$ column equal to $\Sigma_{i-j,i-l}$. In this problem, the complete and incomplete data at time m are denoted by $\mathcal{C}_m = \{\mathcal{Y}_m, \hbar_m, \mathcal{X}_m\}$ and $\mathcal{I}_m = \{\mathcal{Y}_m\}$, respectively, and $\theta_{m|m} = \{\mathcal{U}_{m|m}, \mathbf{C}\}$. The K-L measure with respect to $\mathcal{U}_{m|m}$ in this case is given by [6]

$$\begin{aligned} &Q_m(\mathcal{U}_{m|m}, \hat{\mathbf{C}}_{|m-1}|\hat{\theta}^{(l-1)}_{m|m}) \\ &= \mathrm{E}\{\log p(\mathcal{Y}_m, \mathcal{X}_m, \hbar_m|\mathcal{U}_{m|m}, \hat{\mathbf{C}}_{|m-1})|\mathcal{Y}_m, \hat{\theta}^{(l-1)}_{m|m}\} \\ &= \textstyle\sum_{i=1}^{m}\{-\rho_i(\tilde{\mu}_{i|m}, \mu_{i|m}) - \eta_i(\mu_{i|m}, \bar{\mu}_{i|m}) + c_i\} + d \\ &-[\mu_{0|m} - \mu_{-1}]^H \Sigma_{-1}^{-1}[\mu_{0|m} - \mu_{-1}] - \rho_0(\mu_{0|m}, \tilde{\mu}_{0,m}), \end{aligned} \quad (7)$$

where c_i and d are constant terms, and

$$\begin{aligned} \rho_i(\tilde{\mu}_{i|m}, \mu_{i|m}) &= [\tilde{\mu}_{i|m} - \mu_{i|m}]^H \tilde{\mathbf{C}}_{i|m}[\tilde{\mu}_{i|m} - \mu_{i|m}], \\ \eta_i(\mu_{i|m}, \bar{\mu}_{i|m}) &= [\mu_{i|m} - \bar{\mu}_{i|m}]^H (\mathbf{B}\mathbf{B}^H)^{-1}[\mu_{i|m} - \bar{\mu}_{i|m}], \end{aligned}$$

and $\bar{\mu}_{i|m} = \check{\mathbf{F}}\underline{\mu}_{i-1|m}$, $\check{\mathbf{F}} = [\mathbf{F}_1, \cdots, \mathbf{F}_{L_h}]$ and μ_{-1} and Σ_{-1} are initial value for recursion setup. $\tilde{\mathbf{C}}_{i|m}$ and $\tilde{\mu}_{i|m}$ come from the synthetic system that can be characterized by

$$\begin{aligned} \tilde{\mathbf{C}}_{i|m} &= \mathrm{E}\{\mathbf{T}_i^H \hat{\mathbf{C}}_{|m-1}^{-1}\mathbf{T}_i|\mathcal{Y}_m, \hat{\theta}^{(l-1)}_{m|m}\}, \\ \tilde{\mathbf{T}}_{i|m} &= \mathrm{E}\{\hat{\mathbf{C}}_{|m-1}^{-1}\mathbf{T}_i|\mathcal{Y}_m, \hat{\theta}^{(l-1)}_{m|m}\}, \end{aligned}$$

with the synthetic log likelihood function of $\{\mathbf{h}_i, \hat{\mathbf{C}}_{|m-1}\}$ at time i defined as

$$\begin{aligned} &SLL_i(\mathbf{h}_i, \hat{\mathbf{C}}_{|m-1}|\mathcal{Y}_m, \hat{\theta}^{(l-1)}_{m|m}) = -\log(\pi^{L_r}\det(\hat{\mathbf{C}}_{|m-1})) \\ &-\mathrm{E}\{(\mathbf{y}_i - \mathbf{T}_i\mathbf{h}_i)^H \hat{\mathbf{C}}_{|m-1}^{-1}(\mathbf{y}_i - \mathbf{T}_i\mathbf{h}_i)|\mathcal{Y}_m, \hat{\theta}^{(l-1)}_{m|m}\} \\ &= -[\tilde{\mathbf{C}}_{i|m}^{-1}\tilde{\mathbf{T}}_{i|m}\mathbf{y}_i - \mathbf{h}_i]^H \tilde{\mathbf{C}}_{i|m}^{-1}[\tilde{\mathbf{C}}_{i|m}^{-1}\tilde{\mathbf{T}}_{i|m}\mathbf{y}_i - \mathbf{h}_i] + const., \end{aligned}$$

which can be maximized (with respect to $\mathbf{h}_i$) by setting $\tilde{\mu}_{i|m} = \tilde{\mathbf{C}}_{i|m}^{-1}\tilde{\mathbf{T}}_{i|m}\mathbf{y}_i$. Therefore, $\tilde{\mu}_{i|m}$ is referred to as the synthetic ML estimate of $\mathbf{h}_i$. Notice that, in (7), except for the constant terms, all other terms are in a Gaussian quadratic form. Thus, maximizing (7) with respect to $\mathcal{U}_{m|m}$ is equivalent to finding

$$\begin{aligned} \hat{\mathcal{U}}_{m|m} &= \arg\max_{\mathcal{U}_{m|m}} P(\mu_{0|m}|\mu_{-1})P(\tilde{\mu}_{0|m}|\mu_{0|m}) \\ &\quad \prod_{i=1}^{m} P(\tilde{\mu}_{i|m}|\mu_{i|m})P(\mu_{i|m}|\bar{\mu}_{i|m}). \end{aligned} \quad (8)$$

This is also equivalent to seeking the recursive expression of $\hat{\mathcal{U}}_{i|m}$. The probabilistic model implies that the time evolution characteristic of $\mathcal{U}_{i|m}$ can be modeled by

$$\begin{aligned} \mu_{i|m} &= \check{\mathbf{F}}\underline{\mu}_{i-1|m} + \mathbf{B}\nu_i, \\ \tilde{\mu}_{i|m} &= \mu_{i|m} + \omega_i, \end{aligned}$$

where $\nu_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\omega_i \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{C}}_{i|m}^{-1})$. The recursive stochastic MMSE channel estimator is given by a series of Kalman-like equations. They can be written as

$$\begin{aligned} \hat{\mathcal{U}}_{i|m} &= \hat{\mathcal{U}}'_{i|m} + \mathbf{K}_{i|m}(\tilde{\mu}_{i|m} - \mathbf{J}_i^H \hat{\mathcal{U}}'_{i|m}), && (9) \\ \hat{\Gamma}_{i|m} &= \hat{\Gamma}'_{i|m} - \mathbf{K}_{i|m}\mathbf{J}_i^H \hat{\Gamma}'_{i|m}, && (10) \\ \mathbf{K}_{i|m} &= \hat{\Gamma}'_{i|m}\mathbf{J}_i(\tilde{\mathbf{C}}_{i|m}^{-1} + \mathbf{J}_i^H \hat{\Gamma}'_{i|m}\mathbf{J}_i)^{-1}, && (11) \end{aligned}$$

where $\mathbf{J}_i = [\mathbf{I}_{L_tL_p \times L_tL_p}, \mathbf{0}_{L_tL_p \times iL_tL_p}]^H$ and $\hat{\mathcal{U}}'_{i|m}$ and $\hat{\Gamma}'_{i|m}$ are obtained via

$$\begin{aligned} \hat{\mathcal{U}}'_{i|m} &= [\bar{\mu}_{i|m}^T, \mathcal{U}_{i-1|m}^T]^T = [(\check{\mathbf{F}}\underline{\mu}_{i-1|m})^T, \mathcal{U}_{i-1|m}^T]^T, \\ \hat{\Gamma}'_{i|m} &= \begin{bmatrix} \check{\mathbf{F}}\underline{\hat{\Sigma}}_{i-1,i-1|m}\check{\mathbf{F}}^H & \check{\mathbf{F}}\mathbf{V}_{i-1|m}^H \\ \mathbf{V}_{i-1|m}\check{\mathbf{F}}^H & \hat{\Gamma}_{m-1|m} \end{bmatrix} + \mathbf{J}_i\mathbf{B}\mathbf{B}^H\mathbf{J}_i^H, \end{aligned}$$

where $\mathbf{V}_{i-1|m} = \mathrm{E}\{[\mathcal{U}_{i-1|m} - \hat{\mathcal{U}}_{i-1|m}][\underline{\mu}_{i-1|m} - \underline{\hat{\mu}}_{i-1|m}]^H\}$. For every time index i, the estimates from time up to i get updated. The dimension of $\hat{\Gamma}_{i|m}$ also increases by L_tL_p at each time step. However, the dimension of matrix $\tilde{\mathbf{C}}_{i|m}^{-1} + \mathbf{J}_i^H \hat{\Gamma}'_{i|m}\mathbf{J}_i$, which requires inversion, is still $L_tL_p \times L_tL_p$. Although the recursive process is similar to the traditional RLS algorithm or Kalman filtering, the main difference is that a synthetic approach is used to average over the *a posterior* probabilities of $\mathbf{x}$. Therefore, the recursive EM can be viewed as a Kalman-filter like algorithm with soft decision feedback. The sequential update of the MMSE covariance estimator can be approximated by

$$\begin{aligned} \hat{\mathbf{C}}_{|m} &= (1 - \frac{1}{m})\hat{\mathbf{C}}_{|m-1} + \frac{1}{m}\tilde{\mathbf{C}}_m^{MMSE}, && (12) \\ \tilde{\mathbf{C}}_m^{MMSE} &= [\mathbf{y}_m - \hat{\mathbf{T}}_{m|m}\hat{\mu}_{m|m}][\mathbf{y}_m - \hat{\mathbf{T}}_{m|m}\hat{\mu}_{m|m}]^H \\ &\quad - \hat{\mathbf{T}}_{m|m}\hat{\mu}_{m|m}\hat{\mu}_{m|m}^H\hat{\mathbf{T}}_{m|m}^H + \tilde{\mathbf{R}}_{m|m}, && (13) \end{aligned}$$

where $\hat{\mathbf{T}}_{m|m} = \mathrm{E}\{\mathbf{T}_m|\mathcal{Y}_m, \theta^{(l-1)}_{m|m}\}$ and

$$\tilde{\mathbf{R}}_{m|m} = \mathrm{E}\{\mathbf{T}_m\hat{\mu}_{m|m}\hat{\mu}_{m|m}^H\mathbf{T}_m^H|\mathcal{Y}_m; \hat{\theta}_{m|m-1}\}.$$

We summarize the sequential MMSE estimator obtained by EM algorithm by using the Forney-style Factor Graph representation [8] as shown in Fig. 2.

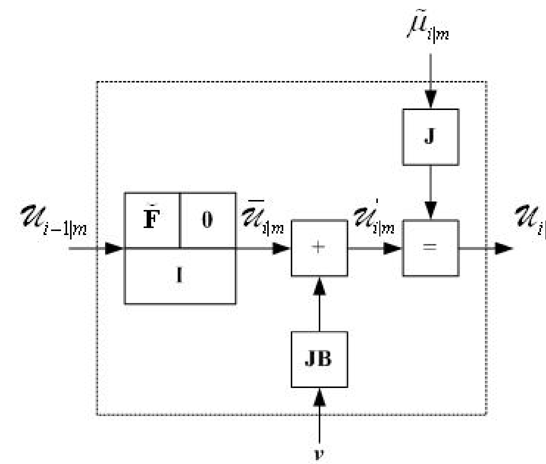

Fig. 2. The Forney-style Factor Graph representation and update rules for the EM estimator by following the notation in Table 1 of [8].

B. Symbol Detection with Soft Decoder

To implement the recursive EM algorithm, the *a posterior* probability (APP) $p(\mathbf{x}_i|\mathcal{Y}_m, \hat{\theta}^{(l-1)}_{m|m})$ is needed. It can be computed by soft decoders. Thus, to implement the joint channel estimation and symbol detection algorithm, the soft decoder and the recursive EM work in concert, with the soft decoder providing the metrics for the EM estimator and the EM providing the likelihoods required for metric updates. The recursive procedure can be represented with a graphical model using the concept of message passing [6].

The BCJR algorithm [9] is probably the most well known soft decoder that provides optimal detection over the inter-symbol interference (ISI) channel. A complete description by a finite state machine at time m would require a trellis diagram with $\mathcal{A}^{(L_c-1)\times L_t} \times m$, where $\mathcal{A}$ is the alpha-beta size of the modulation. For large L_c and L_t, a full state BCJR algorithm seems not feasible. A sliding window BCJR algorithm [10] will lead to suboptimal performance with a much lower complexity by neglecting sequences outside the truncation depth L_d. In order to make detection reliable, we stack all successive received vectors contributed by the transmitted sequence $[x_1(m)\cdots, x_{L_t}(m-L_d+1)]^T$ as

$$\mathbf{z}_m = \begin{bmatrix} \mathbf{H}_m & \mathbf{0}_{L_r\times L_t} & \cdots & \mathbf{0}_{L_r\times L_t} \\ \mathbf{0}_{L_r\times L_t} & \mathbf{H}_{m-1} & \cdots & \mathbf{0}_{L_r\times L_t} \\ \vdots & \mathbf{0}_{L_r\times L_t} & \ddots & \vdots \\ \mathbf{0}_{L_r\times 1} & \cdots & \mathbf{0}_{L_r\times 1} & \mathbf{H}_{m-L_d+1} \end{bmatrix} \mathbf{s}_m + \mathbf{w}_m, \tag{14}$$

where

$$\begin{aligned} \mathbf{z}_m &= [\mathbf{y}_m^T, \cdots, \mathbf{y}_{m-L_d+1}^T]^T, \\ \mathbf{s}_m &= [x_1(m), \cdots, x_{L_t}(m-L_d-L_c+2)]^T, \\ \mathbf{w}_m &= [\mathbf{n}_m^T, \cdots, \mathbf{n}_{m-L_d+1}^T]^T. \end{aligned}$$

Let $\mathbb{X}_{ijk,\pm1} = \{\mathbf{s}|b_i^k(j) = \pm1\}$, where $b_i^k(j)$ is the k-th bit of the symbol $x_i(j)$. Then, the soft information can be derived as

$$\lambda[b_i^k(j)] = \ln\frac{P[b_i^k(j)=1]}{P[b_i^k(j)=1]} + \ln\frac{\sum_{\mathbb{X}_{ijk,+1}} p(\mathbf{z}|\mathbf{s})P(\mathbf{s}|b_i^k(j))}{\sum_{\mathbb{X}_{ijk,+1}} p(\mathbf{z}|\mathbf{s})P(\mathbf{s}|b_i^k(j))}. \tag{15}$$

The MIMO detection is similar to the multi-user detection in CDMA systems with the number of receive antennas equivalent to the processing gain and the number of transmit antennas equivalent to the number of users. The soft decoding procedure applied to CDMA based on sliding window BCJR algorithm has already been proposed in [10], and the approach has been adapted and incorporated into our work.

IV. Simulation Results

The performance of the proposed joint estimation and symbol detection algorithm is studied via computer simulation in this section. The binary-phase-shift-key (BPSK) modulation is adopted with differential encoding/decoding to compensate for the π radian phase ambiguity in channel coefficient estimates. Signals of two transmitters arrive at an uniform linear array of $L_r = 4$ over a time-varying channel of $L_p = 3$ paths. The number of iterations is two. The fading channel is simulated using an ARMA model of order 31 with a normalized Doppler spread $f_D T = 5\times10^{-3}$. The order of the channel model used at the receiver is 3. Thus, there is a mismatch between the model for channel generation and that for channel estimation. Fig. 3 gives the tracking performance for time-varying frequency-selective fading channels . The estimation error becomes slightly larger when deep fading occurs. It demonstrates that the third order AR model used at the receiver is sufficient in capturing most of channel dynamics.

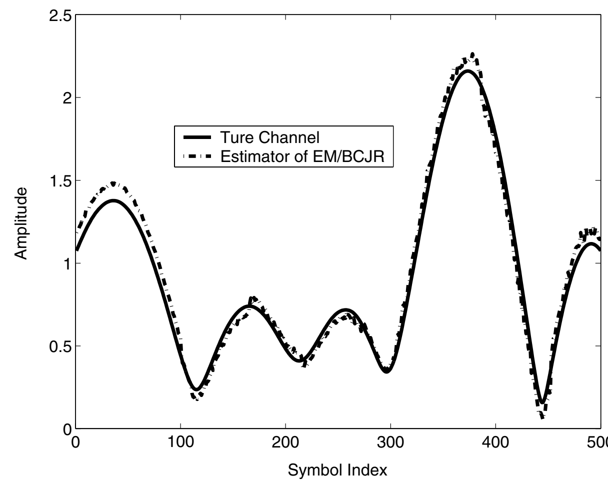

Fig. 3. The tracking performance of the proposed algorithm.

Fig. 4 compares the symbol detection performance using perfect CSI with those using estimates obtained from recursive EM and Kalman filtering with hard decision feedback. The reason for the BER deviation between EM tracking and perfect CSI is the mismatch between the ARMA models for channel generation and estimation. To evaluate the performance of the proposed algorithm, we also consider a receiver employing the Kalman filter with the decision feedback equalizer (DFE). The feedforward and feedback filters of the DFE are chosen to meet the MMSE criteria. The computational complexity for this receiver is much less due to the non-BCJR type detection. However, the price to pay is the inferior performance (about 1.3dB loss for BER at 1×10^{-2}). The relative performance degradation of Kalman filtering with decision feedback to recursive EM is due to the error propagation of the DFE.

In Fig. 5, we examine the performance dependence on ARMA model order selection. The effect of error propagation of the hard decision feedback of Kalman filtering results in an error floor at 5×10^{-3} and 1×10^{-3} for AR(1) and AR(2) models, respectively. One way to avoid this effect is to insert pilot symbols periodically. In contrast, the proposed algorithm does not have an error floor even with low order ARMA model selection.

V. Conclusion

An iterative approach of joint channel tracking and symbol detection algorithm over time-varying MIMO-ISI channels based on the EM algorithm combined with soft decoding was presented. This recursive approach could be viewed as the Kalman filter with soft decision feedback. The performance

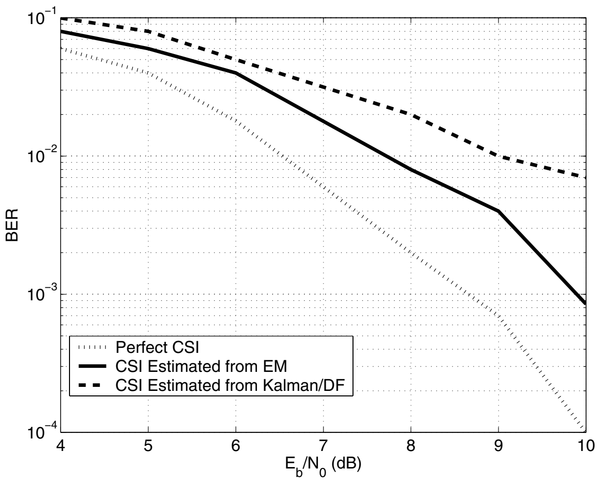

Fig. 4. The bit error rate (BER) performance comparison of three schemes.

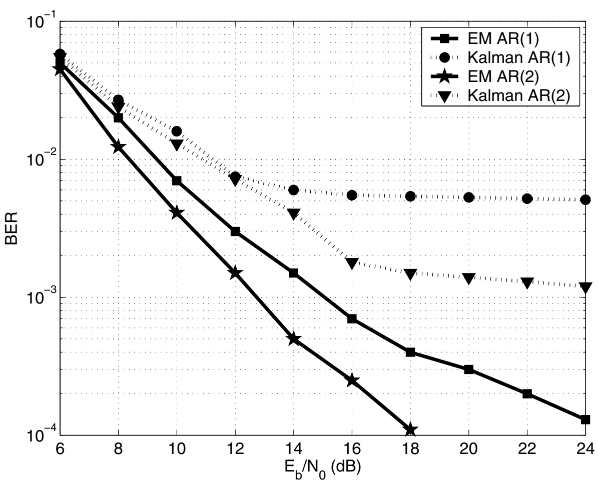

Fig. 5. The bit error rate (BER) performance comparison between the proposed algorithm and Kalman filtering with hard decision feedback with two selected models.

of the proposed algorithm was evaluated via simulations and was shown to be robust to channel model mismatch.

References

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

[2] W. Su, Z. Safar, and K. J. R. Liu, "Space-time signal design for time-correlated Rayleigh fading channels," in *Proc. IEEE ICC*. Anchorage, Alaska, May 2003.

[3] H. Zamiri-Jafarian and S. Pasupathy, "EM-based recursive estimation of channel parameters," *IEEE Trans. on Communications*, vol. 47, no. 9, pp. 1297–1302, Sept. 1999.

[4] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-survivor processing: a general approach to MLSE in uncertain environment," *IEEE Trans. on Communications*, vol. 43, no. 2/3/4, pp. 354–364, 1995.

[5] M. Nissila and S. Pasupathy, "Adaptive Bayesian and EM-based detectors for frequency-selective fading channels," *IEEE Trans. on Communications*, vol. 51, no. 8, pp. 1325–1336, Aug. 2003.

[6] S.-H. Wu, U. Mitra, and C.-C. J. Kuo, "Graph representation for joint channel estimation and symbol detection," in *Proc. IEEE Globecom*. Dallas, Texas, Nov 2004.

[7] M. C. Vanderveen, A. van der Veen, and A. Paulraj, "Estiamtion of multipath parameters in wireless communications," *IEEE Trans. on Signal Processing*, vol. 46, pp. 682–690, Mar. 1998.

[8] H.-A. Loeliger, "Least squares and kalman filtering on forney graphs," in *Codes, Graphs, and Systems, (festchrift in honour of David Forney on the occasion of his 60th birthday), R.E. Blahut and R. Koetter, eds*, 2002, pp. 113–135.

[9] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for mimizing symbol error rate," *IEEE Trans. on Information Theory*, vol. 20, pp. 284–287, Mar. 1974.

[10] S.-H. Wu, U. Mitra, and C.-C. J. Kuo, "Joint channel estimation and multiuser dectection for multipath fading channels in DS-CDMA systems," in *Proc. IEEE ICC*. Anchorage, Alaska, May 2003.