

# Iterative Joint Channel Estimation and Multiuser Detection for DS-CDMA in Frequency-Selective Fading Channels

Sau-Hsuan Wu, *Member, IEEE*, Urbashi Mitra, *Fellow, IEEE*, and C.-C. Jay Kuo, *Fellow, IEEE*

**Abstract**—An iterative joint channel estimation, symbol detection, phase recovery and interference cancellation structure is proposed for asynchronous code-division multiple-access systems over frequency-selective fading channels. Based on the expectation maximization (EM) algorithm, a recursive channel estimator is developed for blind channel tracking, using a novel stochastic signal processing technique. To perform symbol detection given the phase ambiguities of the resultant EM channel estimates, a noncoherent scheme is developed to compute the *a posteriori* probabilities (APPs) of data symbols. Moreover, by incorporating the APPs into the proposed recursive channel estimator, phase ambiguity due to the EM channel estimation can be resolved, which enables soft multiple access interference cancellation for multiuser detection. Based on these new signal processing schemes, an iterative structure is proposed for joint channel estimation and multiuser detection over fast fading channels.

**Index Terms**—DS-CDMA, expectation maximization (EM), joint estimation and detection, multiuser detection, noncoherent detection.

## I. INTRODUCTION

FOR MOBILE wireless communications, a significant amount of training data is often embedded in transmission packets to help estimate rapidly changing channel parameters. For example, one eighth of transmitted symbols are used for training in IEEE 802.16e [1]. This training overhead inevitably consumes a large portion of the effective data bandwidth, thus becoming a major barrier to increasing data throughput in highly dynamic channels. Real-time blind channel tracking

not only eliminates the need of training symbols but also helps reduce estimation errors, which potentially reduces bit error rates and packet dropping rates and, hence, improves data throughput. To overcome the throughput limitation due to training and estimation errors, joint channel estimation and symbol detection in time-varying channels is considered herein.

Over the past decade, much effort has been devoted to developing effective approaches for joint channel estimation and symbol detection (JED) in time-varying channels. Among them, a commonly employed method is pilot-assisted channel estimation, where channel parameters between pilots are interpolated using channel estimates obtained with pilot symbols e.g. [2], [3]. Practical implementations of pilot-assisted channel estimation schemes are often equipped with hard-decision [3], [4] or soft-decision feedback [2], [5]–[7], where “hard-” or “soft-” detected symbols are used to replace unknown transmitted symbols in channel estimation to help reduce channel estimation errors. These approaches are more suitable for quasi-static or slowly fading channels if the number of pilots is not properly matched to the fade rate.

Alternatively, to effectively track fast fading channels, JED schemes have been developed under the assumption of no prior data information for channel estimation. Methods of this kind can be roughly categorized into two classes. Methods in the first class use the minimum mean squared error (MMSE) criterion for channel estimation and the maximum likelihood method for sequence detection (MLSD) based on the per-survivor processing (PSP) technique [8]. Representative work of this class includes [9]–[11] and the references therein. The second class employs probabilistic information about the data symbols to assist channel estimation under the expectation maximization (EM) framework. Pioneering work of this class is the EM-based maximum likelihood (ML) channel estimator in [12] for static channels. Other examples include the maximum *a posteriori* (MAP) estimators in [13], [14] for fading channels.

Both PSP and the methods in [13], [14] can track fast fading channels at a complexity that grows exponentially with the number of resolvable paths in multipath channels. To reduce the complexity, some modifications have been made by trading tracking performance for complexity reduction, e.g. the exponentially complex algorithm and its low-complexity modification in [14]. In the same spirit of the EM-based ML estimation of [12], we develop a recursive estimator for blind channel tracking in frequency-selective time-varying channels, which turns out to be a joint stochastic MMSE channel estimator and MAP detector under additive Gaussian noise. Its

Manuscript received December 4, 2006; revised December 27, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Sergiy A. Vorobyov. This research was supported in part by the following organizations and grants: the National Science Council of Taiwan under Grant NSC 95-2219-E-009-011, the National Science Foundation under Grant NSF/ANI-0087761, and the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, under Cooperative Agreement No. EEC-9529152. This paper was presented in part at the IEEE International Conference on Communications, Anchorage, AK, May 2003, in part at the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Nov. 2004, and at the IEEE Global Telecommunications Conference, Dallas, TX, Dec. 2004.

S.-H. Wu is with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C. (e-mail: sahsuan@cm.nctu.edu.tw).

U. Mitra and C.-C. J. Kuo are with the Ming-Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: ubli@usc.edu; cckuo@sipi.usc.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2008.920146

complexity is much lower than the method presented in [13], [14] with no compromise in its performance. Some preliminary results in this work have been reported in [15]–[17], and a generic graphical representation of the proposed algorithm was presented in [18].

Despite the different criteria employed in symbol detection, the two classes above of JED schemes were found in [10] to fall within a common EM framework [19]. Different schemes result from different choices of the hidden variables in the corresponding EM settings. More specifically, PSP models channel parameters as hidden variables and resolves the joint optimization problem with the Viterbi algorithm, using a Kalman filter for each survival path to track all possible channel realizations. Methods in the second class, model the data symbols as hidden variables and use the BCJR algorithm [20] to generate the probabilistic information required for channel estimation. It was shown in [15]–[18] that the latter approach demands a lower complexity than its PSP counterpart via proper signal processing in the EM procedure.

In addition to narrowband communications, the EM algorithm has also found many applications in joint channel estimation and multiuser detection (MUD) for direct-sequence (DS) code-division multiple access (CDMA) systems. For instance, the ML MUD schemes presented in [21] and [22] for flat fading channels employ the “SAGE” algorithm in [23]. A sequential joint estimator and detector for multipath fading channels was proposed in [24] using hard decision feedback and a multiuser signal decomposition method [25]. An application of [12] was also employed in [26] for MUD in multipath channels with complexity that grows exponentially with the number of users. Despite the rich research results in this area, an effective iterative structure for soft information exchanges between demodulation modules of channel estimation, symbol detection and interference cancellation for multipath fast fading channels is still lacking. This type of structure was shown [27] to be the key ingredient for achieving single-user performance without resorting to exponentially complex MUD algorithms when perfect channel state information (CSI) is available. With this observation, we develop a low-complexity soft iteration structure for joint channel estimation and MUD for DS-CDMA in frequency-selective fast fading channels.

In contrast to previous work on JED for DS-CDMA systems, we propose an iterative soft information exchange mechanism in this work to perform joint MMSE channel estimation, noncoherent MAP detection, phase recovery and multiple access interference (MAI) cancellation in fast fading channels. First, a low-complexity recursive MMSE estimator is developed for blind channel tracking via a novel stochastic signal processing technique derived from the recursive EM algorithm [28]. Next, to perform symbol detection under the phase ambiguities associated with the resulting EM channel estimates [29], a forward-backward message passing scheme is provided to compute the noncoherent *a posteriori* probability (APP) in multipath fading channels, based on the noncoherent BCJR algorithm originally developed in [30] for additive white Gaussian (AWGN) channels. Furthermore, to suppress MAI for the desired user, a channel adjustment method is proposed to compensate for the phase errors of the EM channel estimates by in-

corporating the noncoherent APPs of data symbols into the stochastic signal processing procedure for blind channel tracking.<sup>1</sup> This phase information is necessary in estimating MAI, which together with EM channel estimates and noncoherent APPs not only enables soft MAI cancellation in MUD, but also makes the proposed structure extensible to iterative (Turbo) MAI cancellation and channel decoding, which was shown in [27] to be an effective low-complexity approach for achieving single-user performance in DS-CDMA, given perfect CSI.

The rest of this paper is organized as follows. The system model for multipath DS-CDMA channels is described in Section II. A brief review for the EM algorithm and its application on JED is given in Section III. Based on the recursive EM algorithm, a stochastic recursive estimator is developed in Section IV for blind channel tracking in multipath fading channels. To perform symbol detection under the phase ambiguities resulting from the EM channel estimates, a modified noncoherent MAP detector is derived in Section V for robust MUD. Incorporating the posterior probabilities into the EM channel estimator, a complete signal processing procedure for joint channel tracking, symbol detection and phase error correction is specified in Section VI. Finally, an iterative structure is proposed in Section VII for joint channel estimation, interference cancellation and MUD. Simulation results are presented in Section VIII to demonstrate the performance of the proposed iterative structure, which is followed by concluding remarks in Section IX.

## II. SYSTEM MODEL

Consider an asynchronous DS-CDMA system with  $K$  users. The baseband representation of the transmitted signal for the  $k$ th user can be written as

$$x_k(t) = \sqrt{2P_k} \sum_{n=-\infty}^{\infty} b_k(n) s_k(t - nT_s - \tau_k) \quad (1)$$

where  $T_s$  is the symbol duration, and  $b_k(n)$ ,  $P_k$  and  $\tau_k$  are the data bit at time  $n$ , energy per bit and relative transmission delay with reference to the base station, respectively, for the  $k$ th user. The transmitted symbols  $b_k(n)$  are identically and independently distributed (i.i.d.) random variables taking values from the finite alphabet set of MPSK, i.e.  $b_k(n) \in \{e^{j2\pi m/M_s} : m = 0, 1, \dots, M_s - 1\}$ . The spreading waveform is given by  $s_k(t) = \sum_{i=0}^{N-1} c_k(i) \psi(t - iT_c)$ , where  $c_k(i) \in \{(-1/\sqrt{N}), (1/\sqrt{N})\}$  is the signature sequence of user  $k$ , with a period of  $N$ . The function  $\psi(t)$  is a normalized chip pulse shaping function of duration  $T_c$ , with the spreading gain  $N = T_s/T_c$ .

The  $k$ th user's signal  $x_k(t)$  propagates through a multipath fading channel with the complex impulse response  $f_k(t) = \sum_{l=1}^{L_k} f_{kl}(t) \delta_D(t - \eta_{kl})$ , where  $\delta_D(t)$  is the Dirac delta function,  $L_k$  is the number of paths of user  $k$ ,  $\eta_{kl}$  is the time delay associated with the  $l$ th tap of the tapped-delay-line channel model and  $f_{kl}(t)$  is the fading process corresponding to the tap. The

<sup>1</sup>We note that after incorporating noncoherent APPs, the original blind channel tracking scheme could be viewed as semi-blind when used in estimating MAI, as some pilot symbols are required to serve as the boundary conditions for the noncoherent BCJR algorithm. Nevertheless, no training symbol is needed for the desired user's channel tracking.

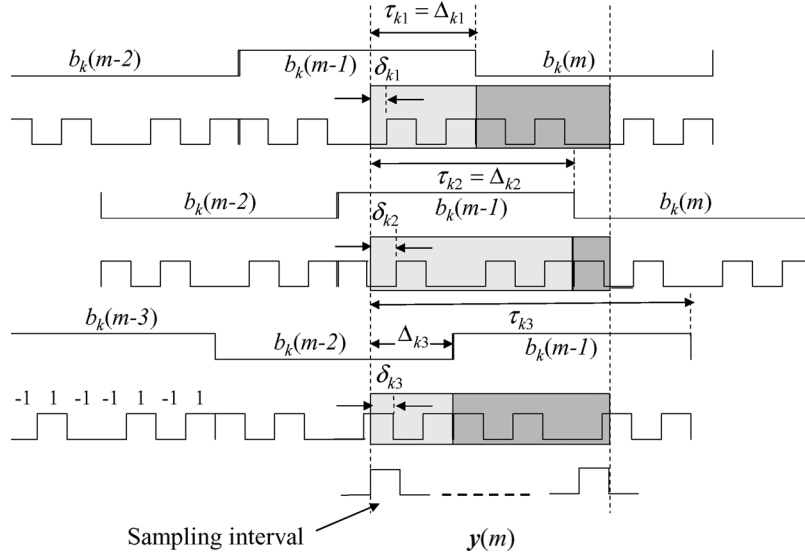


Fig. 1. The delay pattern of a three-path DS-CDMA channel for user  $k$ , where the light-shaded regions belong to symbols that arrive earlier inside the sampling interval of  $\mathbf{y}(m)$  and the dark-shaded regions belong to symbols that arrive later in the same transmission paths of the same sampling interval.

value of  $f_{kl}(t)$  is assumed to be constant during one symbol interval and changes from symbol to symbol. Thus,  $f_{kl}(t)$  can be modeled by a discrete-time fading process  $a_{kl}\lambda_{kl}(m)$ , where  $a_{kl}$  is a time-invariant nonnegative channel gain and  $\lambda_{kl}(m)$  is a complex zero-mean wide sense stationary Gaussian process satisfying  $E[\lambda_{k_1 l_1}^*(m)\lambda_{k_2 l_2}(m)] = \delta_K(k_1 - k_2)\delta_K(l_1 - l_2)$ , where  $\delta_K$  stands for the Kronecker delta function. The autocorrelation between two consecutive channel states is given by  $\rho = E[\lambda_{k_1 l_1}^*(m)\lambda_{k_1 l_1}(m-1)]$ ,  $0 \leq |\rho| < 1$ .

The received signal due to the  $k$ th user is given by

$$\begin{aligned} r_k(t) &= x_k(t) * f_k(t) \\ &= \sqrt{2P_k} \sum_{n=-\infty}^{\infty} b_k(n) \sum_{j=0}^{N-1} \sum_{l=1}^{L_k} f_{kl}(t)c_k(j)\psi \\ &\quad \times (t - nT_s - jT_c - \tau_{kl}) \end{aligned} \quad (2)$$

where  $\tau_{kl} = \tau_k + \eta_{kl}$  is the path delay along the path  $l$  of user  $k$ . The overall received signal is equal to  $r(t) = \sum_{k=1}^K r_k(t) + n(t)$ , where  $K$  is the number of users and  $n(t)$  is a complex zero-mean additive white Gaussian noise (AWGN). Define  $m_{kl} \triangleq \lceil \tau_{kl}/T_s \rceil$ , where  $\lceil \cdot \rceil$  denotes the largest integer less than or equal to the argument,  $\Delta_{kl} \triangleq T_s \lfloor \tau_{kl} \rfloor$ ,  $i_{kl} \triangleq \lfloor \Delta_{kl}/T_c \rfloor$ ,  $0 \leq i_{kl} < N - 1$  and  $\delta_{kl}T_c \triangleq T_c \lfloor \Delta_{kl} \rfloor$ ,  $0 \leq \delta_{kl} < 1$ , where  $a \lfloor b \rfloor$  denotes the remainder of  $b$  divided by  $a$ . An example of a three-path delay pattern for user  $k$  is illustrated in Fig. 1.

The received signal is passed through a filter matched to the chip pulse shaping function  $\psi(t - mT_s - iT_c)$  and then sampled at the chip rate. We define  $\mathbf{c}_k^L(i_{kl}) \triangleq [0 \cdots 0, c_k(0) \cdots c_k(N - i_{kl} - 1)]^T$  and  $\mathbf{c}_k^R(i_{kl}) \triangleq [c_k(N - i_{kl}) \cdots c_k(N - 1), 0 \cdots 0]^T$ ; both vectors are of dimension  $N \times 1$ . Due to the asynchronous arrival times, the contribution of the signature sequence due to the earlier arrival symbol for path  $l$  in Fig. 1 can be expressed as  $\mathbf{s}_{kl}^- \triangleq (1 - \delta_{kl})\mathbf{c}_k^R(i_{kl}) + \delta_{kl}\mathbf{c}_k^R(i_{kl} + 1)$ , and the contribution of

the signature sequence due to the subsequent symbol on path  $l$  is equal to  $\mathbf{s}_{kl}^+ \triangleq (1 - \delta_{kl})\mathbf{c}_k^L(i_{kl}) + \delta_{kl}\mathbf{c}_k^L(i_{kl} + 1)$ . By collecting  $N$  samples of  $y_k$ ,  $[y_k(mN) \cdots y_k(mN + N - 1)]$ , the discrete time received signal vector due to the  $k$ th user is equal to

$$\sum_{l=1}^{L_k} h_{kl}(m) \cdot [\mathbf{s}_{kl}^+ b_k(m - m_{kl}) + \mathbf{s}_{kl}^- b_k(m - m_{kl} - 1)] \quad (3)$$

where  $h_{kl}(m) \triangleq \sqrt{2P_k} a_{kl} \lambda_{kl}(m)$  and  $\mathbf{g}_{kl}(m - m_{kl}) \triangleq \mathbf{s}_{kl}^+ b_k(m - m_{kl}) + \mathbf{s}_{kl}^- b_k(m - m_{kl} - 1)$ . The discrete time received signal can be expressed as

$$\mathbf{y}(m) = \sum_{k=1}^K \sum_{l=1}^{L_k} \mathbf{g}_{kl}(m - m_{kl}) h_{kl}(m) + \mathbf{n}(m). \quad (4)$$

It is clear from this expression that there are at most  $2L_k$  symbols of user  $k$  involved in the received signal  $\mathbf{y}(m)$  due to the delays of multiple incoming paths. Let  $L_k^p \triangleq \lceil \max(\tau_{kl})/T_s \rceil$  be the longest symbol delay for user  $k$ . Then, without loss of generality,  $\mathbf{y}(m)$  can be considered as a function of  $\xi_{k:m} \triangleq [b_k(m), \dots, b_k(m - L_k^p)] \triangleq [b_k(m), \mathbf{v}_{k:m}]$ , as is  $\mathbf{G}_k(\xi_{k:m}) \triangleq [\mathbf{g}_{k1}(m), \dots, \mathbf{g}_{kL_k}(m - L_k^p + 1)]$ . Thus, defining  $\mathbf{h}_k(m) \triangleq [h_{k1}(m), \dots, h_{kL_k}(m)]^T$ , the received signal can be expressed in matrix form as

$$\mathbf{y}(m) = \sum_{k=1}^K \mathbf{G}_k(\xi_{k:m}) \mathbf{h}_k(m) + \mathbf{n}(m). \quad (5)$$

We note that the spreading sequence  $c_k$  and the time delay  $\tau_{kl}$  be fixed and given.<sup>2</sup> The only unknown in the system matrix  $\mathbf{G}_k(\xi_{k:m})$  is then the symbol state  $\xi_{k:m}$ .

<sup>2</sup>The path delay time varies much more slowly than the fading gain, and is usually determined by code and timing acquisition circuits. Therefore, we assume they are fixed and given throughout this work.

### III. JOINT CHANNEL ESTIMATION AND SYMBOL DETECTION USING THE EM ALGORITHM

The complexity of performing channel estimation and symbol detection simultaneously for all users is prohibitively high. The complexity is proportional to  $M_s^{K(L_p+1)}$ , where  $L_p \triangleq \max_k \{L_k^p\}$  is the largest symbol delay among all users. In addition, a large number of samples of the received signal have to be collected to maintain the quality of channel estimates, which is counter to tracking a fast fading channel. To control the algorithm complexity, as well as to reduce the number of signal samples, estimation and detection is performed separately for each user. When dealing with the  $k$ th user's signal, all interfering users' signals are lumped together as an interference vector  $\mathbf{i}_k(m)$ . Therefore, the received signal (5) is rewritten as

$$\begin{aligned} \mathbf{y}(m) &= \mathbf{G}_k(\xi_{k:m})\mathbf{h}_k(m) + \sum_{i=0, i \neq k}^{K-1} \mathbf{G}_i(\xi_{i:m})\mathbf{h}_i(m) + \mathbf{n}(m) \\ &\triangleq \mathbf{G}_k(\xi_{k:m})\mathbf{h}_k(m) + \mathbf{i}_k(m) + \mathbf{n}(m) \end{aligned} \quad (6)$$

where the covariance matrix of  $\mathbf{i}_k(m) + \mathbf{n}(m)$  is  $\mathbf{C}_k = E\{[\mathbf{i}_k(m) + \mathbf{n}(m)][\mathbf{i}_k(m) + \mathbf{n}(m)]^H\}$ .

For the ease of mathematical manipulations, the distribution of  $\mathbf{i}_k(m)$  is approximated as a zero-mean, colored, complex Gaussian random vector with a covariance matrix  $\mathbf{C}_k$ .<sup>3</sup> To improve the performance of single-user channel estimation and symbol detection in multiple access systems, the contribution of  $\mathbf{i}_k(m)$  can be mitigated iteratively through interference cancellation techniques such as serial interference cancellation (SIC) or parallel interference cancellation (PIC) [31] when all users' channel state and symbol information are available at the end of each iteration. To facilitate the exposition of joint estimation and detection, interference cancellation will be addressed at the end of Section V.

Momentarily ignoring interference cancellation, the multiuser detection problem is that of joint estimation of  $\mathbf{h}_k(m)$  and detection of  $b_k(m)$  for each individual user under the assumption of the colored Gaussian noise,  $\mathbf{i}_k(m)$ . However, the complexity of this JED problem is still high. Nevertheless, this complexity issue is ameliorated via iterative optimization schemes such as the EM algorithm. We next introduce the EM algorithm for JED under colored Gaussian MAI.

#### A. The EM Algorithm

Before introducing the EM algorithm, we first define the objectives of channel estimation and symbol detection. We let the set of the unknown channel parameters up to time  $M$  of user  $k$  be  $\Theta_{k:M} = [\mathbf{C}_k, \mathbf{h}_{k:M}, \dots, \mathbf{h}_{k:1}]$  and the corresponding symbol stream be  $\Psi_{k:M} \triangleq \{b_{k:M}, \dots, b_{k:1}\}$ , where, for simplicity of notation,  $\mathbf{h}_k(m)$ ,  $b_k(m)$  and  $\mathbf{y}(m)$  are also denoted by  $\mathbf{h}_{k:m}$ ,

<sup>3</sup>The exact distribution of  $\mathbf{i}_k(m)$  before interference cancellation is in fact a mixture distribution, while an exact modeling for the distribution of  $\mathbf{i}_k(m)$  is not available when considering the residual interference after iterative soft interference cancellation. Simulation results in Section VIII show that the Gaussian assumption does not seem to affect the performance of MUD given CSI, while its influence on the performance of JED will require more theoretical justification.

$b_{k:m}$  and  $\mathbf{y}_m$ , respectively. It is clear that the complete data for estimating the parameter set  $\Theta_{k:M}$  is  $\{\mathbf{Y}_M, \Psi_{k:M}\}$ , where  $\mathbf{Y}_M \triangleq \{\mathbf{y}_M, \dots, \mathbf{y}_1\}$ . However, due to the fact that  $\Psi_{k:M}$  is not, in fact, observed, the *log likelihood* (LLK) of  $\Theta_{k:M}$  becomes

$$\begin{aligned} \mathcal{L}\mathcal{L}\mathcal{K}(\Theta_{k:M}) &\triangleq \log P(\mathbf{Y}_M; \Theta_{k:M}) \\ &= \log \sum_{\{\Psi_{k:M}\}} P(\mathbf{Y}_M | \Psi_{k:M}; \Theta_{k:M}) \\ &\quad \times P(\Psi_{k:M}). \end{aligned} \quad (7)$$

It is, in general, difficult to estimate  $\Theta_{k:M}$  directly from this LLK due to the exponential complexity involved in the exhaustive search for the optimal  $\Psi_{k:M}$ . To reduce the complexity, we use the EM method in [12] to approach this goal iteratively. In contrast to PSP [8], this method essentially performs the ML estimate of  $\Theta_{k:M}$  and the MAP detection of  $\Psi_{k:M}$  in each iteration.

We define the Kullback–Leibler (K-L) measure of  $\Theta_{k:M}$  at iteration  $\ell$  to be

$$\begin{aligned} Q_M \left( \Theta_{k:M} | \hat{\Theta}_{k:M}^{\ell-1} \right) &\triangleq E_{\Psi_{k:M}} \left\{ \log \{ P(\mathbf{Y}_M, \Psi_{k:M}; \Theta_{k:M}) \} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \end{aligned} \quad (8)$$

where  $M$  in  $Q_M(\cdot)$  implies the use of information collected up to time  $M$  and  $E_{\Psi_{k:M}} \{ \cdot | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \}$  denotes the expectation with respect to (w.r.t.) the hidden state  $\Psi_{k:M}$ , using  $P(\Psi_{k:M} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1})$ , which is, in turn, evaluated based upon the estimate,  $\hat{\Theta}_{k:M}^{\ell-1}$ , of  $\Theta_{k:M}$  at iteration  $\ell - 1$ . The EM algorithm, under this setting, performs the ML estimation of  $\Theta_{k:M}$  in two steps as follows:

**E-step:** Compute  $Q_M(\Theta_{k:M} | \hat{\Theta}_{k:M}^{\ell-1})$ ;

**M-step:**  $\hat{\Theta}_{k:M}^\ell = \arg \max_{\Theta_{k:M}} Q_M(\Theta_{k:M} | \hat{\Theta}_{k:M}^{\ell-1})$ .

Obviously, this algorithm is iterative in nature. Given  $\hat{\Theta}_{k:M}^{\ell-1}$ ,  $Q_M(\Theta_{k:M} | \hat{\Theta}_{k:M}^{\ell-1})$  is computed based on the premise of  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1})$  which, in turn, is computed using the BCJR algorithm [20] with the previous estimate  $\hat{\Theta}_{k:M}^{\ell-1}$ .<sup>4</sup> Furthermore, the new estimate  $\hat{\Theta}_{k:M}^\ell$  is obtained by maximizing  $Q_M(\Theta_{k:M} | \hat{\Theta}_{k:M}^{\ell-1})$  w.r.t.  $\Theta_{k:M}$ . Hence, to start the iteration, an initial set  $\hat{\Theta}_{k:M}^0$  must be assigned in advance to compute the  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^0)$ . The initial set can be obtained either with a random guess or from the initial training symbols of a transmission packet. The EM algorithm guarantees  $\mathcal{L}\mathcal{L}\mathcal{K}(\hat{\Theta}_{k:M}^\ell) \geq \mathcal{L}\mathcal{L}\mathcal{K}(\hat{\Theta}_{k:M}^{\ell-1})$ .

Except for some special cases, e.g.  $\mathbf{C}_k = \sigma^2 \mathbf{I}$ , solving the above equations for  $\mathcal{H}_{k:M} \triangleq [\mathbf{h}_{k:M}, \dots, \mathbf{h}_{k:1}]$  and  $\mathbf{C}_k$  simultaneously is, in general, difficult. Nevertheless, this joint optimization problem can be resolved by optimizing for  $\mathcal{H}_{k:M}$  and  $\mathbf{C}_k$  alternatively, using the extended EM framework [10].

Let  $Q_M(\mathcal{H}_{k:M}, \hat{\mathbf{C}}_k^{\ell-1} | \hat{\Theta}_k^{\ell-1})$  be the K-L measure with  $\mathbf{C}_k$  in (8) replaced by  $\hat{\mathbf{C}}_k^{\ell-1}$ . (Note that  $\Theta_{k:M} \triangleq \{\mathcal{H}_{k:M}, \mathbf{C}_k\}$ .) Maximizing the K-L measure w.r.t.  $\mathcal{H}_{k:M}$  gives

$$\hat{\mathcal{H}}_{k:M}^\ell = \arg \max_{\mathcal{H}_{k:M}} Q_M \left( \mathcal{H}_{k:M}, \hat{\mathbf{C}}_k^{\ell-1} | \hat{\Theta}_k^{\ell-1} \right). \quad (9)$$

<sup>4</sup>For details regarding how to use the BCJR algorithm for the computation of the *a posteriori* probabilities in multipath fading channels, one can refer to [15], [27].

Upon the acquisition of  $\hat{\mathcal{H}}_{k:M}^\ell$ , one can further define another K-L measure  $Q_M(\hat{\mathcal{H}}_{k:M}^\ell, \mathbf{C}_k | \hat{\Theta}_k^{\ell-1})$  by substituting  $\hat{\mathcal{H}}_{k:M}^\ell$  for  $\mathcal{H}_{k:M}$  in (8). Similarly, maximizing this K-L measure w.r.t.  $\mathbf{C}_k$  yields

$$\hat{\mathbf{C}}_k^\ell = \arg \max_{\mathbf{C}_k} Q_M \left( \hat{\mathcal{H}}_{k:M}^\ell, \mathbf{C}_k | \hat{\Theta}_k^{\ell-1} \right). \quad (10)$$

Based on [10, Theorem 1], we have  $Q_M(\hat{\Theta}_k^\ell | \hat{\Theta}_k^{\ell-1}) \geq Q_M(\hat{\mathcal{H}}_{k:M}^\ell, \hat{\mathbf{C}}_k^{\ell-1} | \hat{\Theta}_k^{\ell-1}) \geq Q_M(\hat{\Theta}_k^{\ell-1} | \hat{\Theta}_k^{\ell-1})$ , which guarantees that the LLK of  $\hat{\Theta}_k^\ell$  is non-decreasing, namely  $\mathcal{L}\mathcal{L}\mathcal{K}(\hat{\Theta}_k^\ell) \geq \mathcal{L}\mathcal{L}\mathcal{K}(\hat{\Theta}_k^{\ell-1})$ . In the sequel, we will apply this principle for the inference of  $\mathcal{H}_{k:M}$  and  $\mathbf{C}_k$ , alternatively and recursively.

#### IV. STOCHASTIC RECURSIVE ESTIMATION FOR DYNAMIC FADING CHANNELS

The essence of recursive channel tracking lies in the use of a dynamic model to characterize the temporal evolution of the channel parameters. For fading channels, a widely used model is the autoregressive moving average (ARMA) model [32]:

$$\begin{aligned} \mathbf{h}_{k:m} &= [\mathbf{F}_{k:1} \ \mathbf{F}_{k:2} \ \dots \ \mathbf{F}_{k:N_h}] \underline{\mathbf{h}}_{k:m-1} + \mathbf{B}_k \mathbf{v}_{k:m} \\ &\triangleq \underline{\mathbf{F}}_k^H \underline{\mathbf{h}}_{k:m-1} + \mathbf{B}_k \mathbf{v}_{k:m} \end{aligned} \quad (11)$$

where  $\underline{\mathbf{h}}_{k:m} \triangleq [\mathbf{h}_{k:m}^T, \dots, \mathbf{h}_{k:m-N_h+1}^T]^T$ ,  $N_h$  being the order of the channel model. The matrices  $\mathbf{B}_k$  and  $\mathbf{F}_{k:i}$ ,  $i = 1, \dots, N_h$ , are of dimension  $L_k \times L_k$ , and  $\mathbf{v}_{k:m} \sim \mathcal{N}(\mathbf{0}_{L_k \times L_k}, \mathbf{I}_{L_k \times L_k})$ .

From the channel model, it is clear that  $\underline{\mathbf{h}}_{k:m}$  is a time-varying random vector and can be modeled by a hidden Markov model (HMM). Thus, in the absence of both transmission data and CSI, the incomplete data is only the observation to the system, namely  $\mathbf{Y}_M = \{\mathbf{y}_M, \dots, \mathbf{y}_1\}$ , and, as a result, the hidden state for user  $k$  at time  $m$  is redefined for time-varying channels as  $\psi_{k:m} \triangleq \{\xi_{k:m}, \mathbf{h}_{k:m}\}$  and  $\Psi_{k:M} = \{\psi_{k:M}, \dots, \psi_{k:1}\}$ . The corresponding unknown parameter set to be estimated for user  $k$  is also redefined as  $\Theta_{k:M} = \{\mathbf{C}_k, \mu_{k:M}, \dots, \mu_{k:1}\}$ , where  $\mu_{k:m} \triangleq E[\mathbf{h}_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1}]$ . We note that the expectation is w.r.t. the posterior probability of  $\mathbf{h}_{k:m}$ , given the system parameter  $\hat{\Theta}_{k:M}^{\ell-1}$  of the previous iteration. Furthermore,  $\mathbf{C}_k$  alters only when system loading changes, upon the joining or leaving of users. So, it is still modeled as a constant matrix within each processing block.

Under the above EM setting, the K-L measure can be rewritten in a recursive fashion as

$$\begin{aligned} &Q_M \left( \Theta_{k:M} | \hat{\Theta}_{k:M}^{\ell-1} \right) \\ &= E_{\Psi_{k:M}} \left\{ \log p \left( \mathbf{Y}_M, \Psi_{k:M} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right) \right\} \\ &= E_{\Psi_{k:M}} \left\{ \log p \left( \mathbf{Y}_{m-1}, \Psi_{k:m-1} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right) \right\} \\ &\quad + E_{\Psi_{k:m}} \left\{ \log p(\mathbf{y}_m | \xi_{k:m}, \mathbf{h}_{k:m}) \right. \\ &\quad \quad \left. + \log p(\mathbf{h}_{k:m} | \underline{\mathbf{h}}_{k:m-1}) | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \\ &= Q_{m-1} \left( \Theta_{k:m-1} | \hat{\Theta}_{k:M}^{\ell-1} \right) - \log \{ \pi^N \det(\mathbf{C}_k) \} \\ &\quad - \log \{ \pi^{L_k} \det(\mathbf{B}_k \mathbf{B}_k^H) \} \end{aligned}$$

$$\begin{aligned} &- E_{\psi_{k:m}} \left\{ [\mathbf{y}_m - \mathbf{G}_k(\xi_{k:m}) \mathbf{h}_{k:m}]^H \mathbf{C}_k^{-1} \right. \\ &\quad \left. \times [\mathbf{y}_m - \mathbf{G}_k(\xi_{k:m}) \mathbf{h}_{k:m}] | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \\ &- E_{\Psi_{k:m}} \left\{ [\mathbf{h}_{k:m} - \underline{\mathbf{F}}_k^H \underline{\mathbf{h}}_{k:m-1}]^H (\mathbf{B}_k \mathbf{B}_k^H)^{-1} \right. \\ &\quad \left. \times [\mathbf{h}_{k:m} - \underline{\mathbf{F}}_k^H \underline{\mathbf{h}}_{k:m-1}] | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \end{aligned} \quad (12)$$

where  $\underline{\mathbf{F}}_k^H \triangleq [\mathbf{F}_{k:1}, \dots, \mathbf{F}_{k:N_h}]$ . Taking expectation w.r.t.  $\mathbf{h}_{k:m}$  and using the fact that  $E\{\mathbf{h}^H \mathbf{A} \mathbf{h}\} = E\{[\mathbf{h} - \mu]^H \mathbf{A} [\mathbf{h} - \mu]\} + \mu^H \mathbf{A} \mu \triangleq \text{Tr}\{\Sigma \mathbf{A}\} + \mu^H \mathbf{A} \mu$ , it can be shown that

$$\begin{aligned} &Q_M \left( \Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1} \right) \\ &= Q_{m-1} \left( \Theta_{k:m-1} | \hat{\Theta}_{k:M}^{\ell-1} \right) - \log \{ \pi^N \det(\mathbf{C}_k) \} \\ &\quad - \log \{ \pi^{L_k} \det(\mathbf{B}_k \mathbf{B}_k^H) \} - \mathbf{y}_m^H \mathbf{C}_k^{-1} \mathbf{y}_m \\ &\quad + \mu_{k:m}^H E_{\xi_{k:m}} \left\{ \mathbf{G}_k(\xi_{k:m})^H \mathbf{C}_k^{-1} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \mathbf{y}_m \\ &\quad + \mathbf{y}_m^H E_{\xi_{k:m}} \left\{ \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_{k:m}) | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \mu_{k:m} \\ &\quad - \mu_{k:m}^H E_{\xi_{k:m}} \left\{ \mathbf{G}_k(\xi_{k:m})^H \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_{k:m}) | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \mu_{k:m} \\ &\quad - D'_m - \left[ \mu_{k:m} - \underline{\mathbf{F}}_k^H \underline{\mu}_{k:m-1} \right]^H \\ &\quad \times (\mathbf{B}_k \mathbf{B}_k^H)^{-1} \left[ \mu_{k:m} - \underline{\mathbf{F}}_k^H \underline{\mu}_{k:m-1} \right] - E'_m \end{aligned} \quad (13)$$

where

$$\underline{\mu}_{k:m} \triangleq E[\underline{\mathbf{h}}_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1}],$$

$$D'_m \triangleq \text{Tr}\{\Sigma_{k:m} E_{\xi_{k:m}} \{ \mathbf{G}_k(\xi_{k:m})^H \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_{k:m}) | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \}\}$$

and

$$\begin{aligned} E'_m &\triangleq \text{Tr}\{\Sigma_{k:m-1} \underline{\mathbf{F}}_k (\mathbf{B}_k \mathbf{B}_k^H)^{-1} \underline{\mathbf{F}}_k^H + \underline{\mathbf{F}}_k (\mathbf{B}_k \mathbf{B}_k^H)^{-1} \check{\Sigma}_{k:m,m-1} \\ &\quad + \check{\Sigma}_{k:m,m-1}^H (\mathbf{B}_k \mathbf{B}_k^H)^{-1} \underline{\mathbf{F}}_k^H\} + \text{Tr}\{\Sigma_{k:m} (\mathbf{B}_k \mathbf{B}_k^H)^{-1}\}. \end{aligned}$$

The corresponding covariance matrices are defined as

$$\begin{aligned} \Sigma_{k:m} &\triangleq E_{\Psi_{k:m}} \{ [\underline{\mathbf{h}}_{k:m} - \underline{\mu}_{k:m}] [\underline{\mathbf{h}}_{k:m} - \underline{\mu}_{k:m}]^H | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \}, \\ \check{\Sigma}_{k:m,m-1} &\triangleq E_{\Psi_{k:m}} \{ [\underline{\mathbf{h}}_{k:m} - \underline{\mu}_{k:m}] [\underline{\mathbf{h}}_{k:m-1} - \underline{\mu}_{k:m-1}]^H | \\ &\quad \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \} \end{aligned}$$

and

$$\Sigma_{k:m} \triangleq E_{\psi_{k:m}} \{ [\mathbf{h}_{k:m} - \mu_{k:m}] [\mathbf{h}_{k:m} - \mu_{k:m}]^H | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \}.$$

Notice that  $\mu_{k:m} \triangleq E[\mathbf{h}_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1}]$  is a time-varying variable, and our objective is to find the ML estimates of  $\mu_{k:m} \triangleq [\mu_{k:m}^H, \dots, \mu_{k:1}^H]^H$  and  $\mathbf{C}_k$ . To this end, we need to evaluate the expectation terms present in (13). Since there are no closed-form expressions for the expectations, for brevity, we define

$$\tilde{\mathbf{C}}_{k:m}(\mathbf{C}_k) \triangleq E_{\xi_{k:m}} \left\{ \mathbf{G}_k^H(\xi_{k:m}) \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_{k:m}) | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \quad (14)$$

$$\tilde{\mathbf{S}}_{k:m}(\mathbf{C}_k) \triangleq \mathbf{C}_k^{-1} E_{\xi_{k:m}} \left\{ \mathbf{G}_k(\xi_{k:m}) | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\}. \quad (15)$$

The computational complexity for evaluating these two terms is  $O(M_s^{L_k} N^3 L_k^3)$ . Again, the APP  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1})$  required for the expectation calculation can be computed with the well-known BCJR algorithm, given  $\mathbf{Y}_M$  and  $\hat{\Theta}_{k:M}^{\ell-1}$  of the previous iteration.

Now, with the introduction of these two matrices, (12) can be reformulated as

$$Q_m \left( \Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1} \right) = Q_{m-1} \left( \Theta_{k:m-1} | \hat{\Theta}_{k:M}^{\ell-1} \right) \\ - (\tilde{\mu}_{k:m} - \mu_{k:m})^H \tilde{\mathbf{C}}_{k:m} (\tilde{\mu}_{k:m} - \mu_{k:m}) \\ - \left( \mu_{k:m} - \mathbf{F}_k^H \underline{\mu}_{k:m-1} \right)^H (\mathbf{B}_k \mathbf{B}_k^H)^{-1} \\ \times \left( \mu_{k:m} - \mathbf{F}_k^H \underline{\mu}_{k:m-1} \right) - F'_m \quad (16)$$

where

$$\tilde{\mu}_{k:m} \triangleq \tilde{\mathbf{C}}_{k:m}^{-1} \tilde{\mathbf{S}}_{k:m}^H \mathbf{y}_m$$

and

$$F'_m \triangleq \mathbf{y}_m^H \mathbf{C}_k^{-1} \mathbf{y}_m - \mathbf{y}_m^H \tilde{\mathbf{S}}_{k:m} \tilde{\mathbf{C}}_{k:m}^{-1} \tilde{\mathbf{S}}_{k:m}^H \mathbf{y}_m \\ + \text{Tr} \{ \tilde{\mathbf{C}}_{k:m} \Sigma_{k:m} \} + \log \{ \pi^N \det(\mathbf{C}_k) \} \\ + \log \{ \pi^{L_k} \det(\mathbf{B}_k \mathbf{B}_k^H) \} + D'_m + E'_m.$$

It is obvious from (16) that the product term of  $\mu_{k:m}$  and  $\tilde{\mathbf{C}}_{k:m}$  makes it difficult to solve for both  $\mathcal{U}_{k:m}$  and  $\mathbf{C}_k$  simultaneously. To get around this difficulty, we apply the iterative optimization procedure stated in (9) and (10). We first perform the maximization w.r.t.  $\mathcal{U}_{k:m}$  in the next section and, hence, set  $\mathbf{C}_k$  to  $\tilde{\mathbf{C}}_k^{\ell-1}$ .

#### A. Recursive EM Estimation for Blind Channel Tracking

By setting  $\mathbf{C}_k$  to  $\tilde{\mathbf{C}}_k^{\ell-1}$ , the K-L measure (16) becomes a Gaussian quadratic form in  $\mathcal{U}_{k:m}$ , given that  $F'_m$  is independent of  $\mathcal{U}_{k:m}$  and can, thus, be ignored during the maximization. The maximization w.r.t.  $\mathcal{U}_{k:m}$  can be stated as

$$\hat{\mathcal{U}}_{k:m}^\ell = \arg \max_{\mathcal{U}_{k:m}} Q_m \left( \Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1} \right). \quad (17)$$

The complexity of directly solving this maximization problem is extremely high due to the growing dimension of  $\mathcal{U}_{k:m}$  with the time index  $m$ . This computational complexity can be alleviated with the recursive maximization procedure proposed in [28], which leads to

$$\hat{\mathcal{U}}_{k:m}^\ell = \hat{\mathcal{U}}_{k:m|m-1}^\ell - \left( \frac{\partial^2 Q_m \left( \Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1} \right)}{\partial \mathcal{U}_{k:m}^* \partial \mathcal{U}_{k:m}^T} \bigg|_{\hat{\mathcal{U}}_{k:m|m-1}^\ell} \right)^{-1} \\ \times \left( \frac{\partial Q_m \left( \Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1} \right)}{\partial \mathcal{U}_{k:m}^*} \bigg|_{\hat{\mathcal{U}}_{k:m|m-1}^\ell} \right) \quad (18)$$

where

$$\hat{\mathcal{U}}_{k:m|m-1}^\ell \triangleq [(\mathbf{F}_k^H \hat{\underline{\mu}}_{k:m-1}^\ell)^H, (\hat{\mathcal{U}}_{k:m-1}^\ell)^H]^H,$$

and by the subscript  $m|m-1$ , it means that the present channel estimate is predicted based on the observation  $\mathbf{Y}_{m-1}$ . This algorithm is often referred to as the recursive EM algorithm.

Despite its lower complexity, we note that the recursive EM algorithm gives the exact solution to (17) due to the fact that  $Q_m(\Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1})$  is a Gaussian quadratic form in  $\mathcal{U}_{k:m}$ . Nevertheless, the computational complexity still increases with time. To maintain a consistent algorithm complexity, it is necessary to constrain the dimension of recursion. Hence, the recursive algorithm is modified slightly into

$$\hat{\underline{\mu}}_{k:m}^\ell = \hat{\underline{\mu}}_{k:m|m-1}^\ell - \left( \frac{\partial^2 Q_m \left( \Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1} \right)}{\partial \underline{\mu}_{k:m}^* \partial \underline{\mu}_{k:m}^T} \bigg|_{\hat{\underline{\mu}}_{k:m|m-1}^\ell} \right)^{-1} \\ \times \left( \frac{\partial Q_m \left( \Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1} \right)}{\partial \underline{\mu}_{k:m}^*} \bigg|_{\hat{\underline{\mu}}_{k:m|m-1}^\ell} \right) \quad (19)$$

where we have defined  $\underline{\mu}_{k:m} \triangleq [\mu_{k:m}^H, \mu_{k:m-1}^H]^H$  and  $\hat{\underline{\mu}}_{k:m|m-1}^\ell \triangleq [(\mathbf{F}_k^H \hat{\underline{\mu}}_{k:m-1}^\ell)^H, (\hat{\underline{\mu}}_{k:m-1}^\ell)^H]^H$ , both of dimension  $(N_h + 1)L_k \times 1$ .

Now, let

$$(\mathbf{P}_m^\ell)^{-1} \triangleq -(\partial^2 Q_m(\Theta_{k:m} | \hat{\Theta}_{k:M}^{\ell-1}) / \partial \underline{\mu}_{k:m}^* \partial \underline{\mu}_{k:m}^T) \big|_{\hat{\underline{\mu}}_{k:m|m-1}^\ell}$$

and

$$(\mathbf{P}_{m-1}^\ell)^{-1} \triangleq -(\partial^2 Q_{m-1}(\Theta_{k:m-1} | \hat{\Theta}_{k:M}^{\ell-1}) / \partial \underline{\mu}_{k:m-1}^* \partial \underline{\mu}_{k:m-1}^T) \big|_{\hat{\underline{\mu}}_{k:m-1}^\ell}.$$

By the Matrix Inversion Lemma, it can be shown from (16) that

$$\mathbf{P}_m^\ell = \left[ \begin{array}{c|c} \tilde{\mathbf{C}}_{k:m}^{-1} & \mathbf{0}_{L_k \times N_h L_k} \\ \hline \mathbf{0}_{N_h L_k \times L_k} & \mathbf{P}_{m-1}^\ell \end{array} \right] - \left[ \begin{array}{c} \tilde{\mathbf{C}}_{k:m}^{-1} \\ \hline -\mathbf{P}_{m-1}^\ell \mathbf{F}_k \end{array} \right] \\ \times \left( \tilde{\mathbf{C}}_{k:m}^{-1} + \mathbf{B}_k \mathbf{B}_k^H + \mathbf{F}_k^H \mathbf{P}_{m-1}^\ell \mathbf{F}_k \right)^{-1} \left[ \tilde{\mathbf{C}}_{k:m}^{-1} - \mathbf{F}_k^H \mathbf{P}_{m-1}^\ell \right]. \quad (20)$$

Substituting (16) and (20) into (19), it is straightforward to show that

$$\hat{\underline{\mu}}_{k:m}^\ell = \hat{\underline{\mu}}_{k:m|m-1}^\ell + \left[ \frac{\mathbf{B}_k \mathbf{B}_k^H + \mathbf{F}_k^H \mathbf{P}_{m-1}^\ell \mathbf{F}_k}{\mathbf{P}_{m-1}^\ell \mathbf{F}_k} \right] \\ \times \left( \tilde{\mathbf{C}}_{k:m}^{-1} + \mathbf{B}_k \mathbf{B}_k^H + \mathbf{F}_k^H \mathbf{P}_{m-1}^\ell \mathbf{F}_k \right)^{-1} \left( \tilde{\mu}_{k:m} - \mathbf{F}_k^H \hat{\underline{\mu}}_{k:m-1}^\ell \right). \quad (21)$$

This is a generalized stochastic Kalman-like recursive filter in the sense that no exact data information is needed for the computation of the channel estimator. With the knowledge of transmitted symbols, the estimator degenerates to the Kalman filter proposed in [28]. However, in the absence of the exact

symbol state  $\xi_{k:m}$ ,  $\tilde{\mathbf{C}}_{k:m}$  and  $\tilde{\mathbf{S}}_{k:m}$  serve as the expected matrices of  $\mathbf{G}_k^H(\xi_{k:m})\mathbf{C}_k^{-1}\mathbf{G}_k(\xi_{k:m})$  and  $\mathbf{C}_k^{-1}\mathbf{G}_k(\xi_{k:m})$ , respectively, using  $P(\xi_{k:m}|\mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1})$  obtained with the BCJR algorithm and  $\hat{\Theta}_{k:M}^{\ell-1}$  of the previous iteration. Moreover, during the recursion process from time  $m-1$  to time  $m$ , not only is the new estimate  $\hat{\mu}_{k:m}$  obtained, the estimate  $\hat{\mu}_{k:m-1}$  also gets updated with the addition of innovation term  $\hat{\mu}_{k:m}$ . As a result, the dimension of  $\mathbf{P}_{m-1}^\ell$  increases from  $N_h L_k \times N_h L_k$  to the dimension  $(N_h+1)L_k \times (N_h+1)L_k$  of  $\mathbf{P}_m^\ell$ . Therefore, to maintain a consistent dimension of (21), only the upper left sub-block matrix of dimension  $N_h L_k \times N_h L_k$  of  $\mathbf{P}_m^\ell$  is kept in (20) as the recursion proceeds from time  $m$  to  $m+1$ .

The algorithm's complexity is dominated by the matrix inversion of  $\tilde{\mathbf{C}}_{k:m}$  of dimension  $L_k \times L_k$  and the computation required for evaluating the matrices,  $\tilde{\mathbf{C}}_{k:m}$  and  $\tilde{\mathbf{S}}_{k:m}$ , which is  $O(M_s^{L_k} N^3 L_k^3)$ . Other than that, its complexity is similar to a standard Kalman filter.

We note that a soft decision-directed (SDD) Kalman filter was also proposed in [14] for frequency selective fading channels, based on the Bayesian EM algorithm [33]. Instead of using the averaged state-space model proposed in [33], the approach in [14] essentially constructs a filter for each possible received signal  $\mathbf{y}(\xi_{k:m})$  such that the SDD Kalman filter requires  $M_s^{L_k+1}$  iterations of filtering to obtain an "averaged" Kalman filtering for each time step. As a result, its complexity is one order higher than that of the PSP-based channel estimator [8] which requires  $M_s^{L_k}$  Kalman filters at each time step, despite the Viterbi algorithm used in PSP for the MLSD of  $\xi_{k:m}$  as opposed to the BCJR algorithm used in [14] for the MAP detection of  $\xi_{k:m}$ . To reduce complexity, [14] also proposed for PSK modulation a reduced complexity SDD (RC-SDD) Kalman filter based on the EM-based decomposition method in [34]. The number of filtering for the RC-SDD Kalman filter can consequently be reduced to  $L_k+1$  at the cost of inferior performance to the SDD Kalman filter.

Starting from a different perspective, in this paper we investigate JED with the original EM algorithm [19]. Through the introduction of the synthetic parameters  $\tilde{\mathbf{C}}_{k:m}$  and  $\tilde{\mathbf{S}}_{k:m}$ , we obtain a generalized Kalman filter in (21) which requires only one iteration of filtering to obtain the estimate of  $\mu_{k:m}$  as opposed to  $M_s^{L_k+1}$  iterations required by the SDD Kalman filter and  $L_k+1$  iterations required by its simplified RC-SDD version in [14].

### B. Recursive Estimation of the Noise Covariance Matrix

Given the new channel update  $\hat{\mathbf{U}}_{k:m}^\ell$ , one can form the K-L measure,  $Q(\hat{\mathbf{U}}_{k:m}^\ell, \mathbf{C}|\hat{\Theta}_{k:M}^{\ell-1})$ , for the estimation of  $\mathbf{C}$ , which can be expressed as

$$\begin{aligned} & Q\left(\hat{\mathbf{U}}_{k:m}^\ell, \mathbf{C}|\hat{\Theta}_{k:M}^{\ell-1}\right) \\ &= -m [N \log \pi + \log \{\det(\mathbf{C})\}] \\ &+ \sum_{i=1}^m E_{\xi_{k:i}} \left\{ -[\mathbf{y}_i - \mathbf{G}(\xi_{k:i})\hat{\mu}_{k:i}^\ell]^H \mathbf{C}^{-1} \right. \\ &\quad \left. \times [\mathbf{y}_i - \mathbf{G}(\xi_{k:i})\hat{\mu}_{k:i}^\ell] | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\}. \quad (22) \end{aligned}$$

Taking its derivative w.r.t.  $\mathbf{C}$  and setting to a zero matrix, the covariance estimate is given by

$$\begin{aligned} \hat{\mathbf{C}}_{k:m} &= \frac{1}{m} \sum_{i=1}^m E_{\xi_{k:i}} \\ &\quad \times \left\{ [\mathbf{y}_i - \mathbf{G}_k(\xi_{k:i})\hat{\mu}_{k:i}^\ell] \right. \\ &\quad \left. \times [\mathbf{y}_i - \mathbf{G}_k(\xi_{k:i})\hat{\mu}_{k:i}^\ell]^H | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \\ &= \left(1 - \frac{N_h+1}{m}\right) \hat{\mathbf{C}}_{k:m-N_h-1} \\ &\quad + \frac{N_h+1}{m} \Delta \tilde{\mathbf{C}}_{k:m} \quad (23) \end{aligned}$$

where  $\hat{\mathbf{C}}_{k:m-N_h-1}$  is the previous estimate at time  $m-N_h-1$ , and the new update term

$$\begin{aligned} \Delta \tilde{\mathbf{C}}_{k:m} &\triangleq \sum_{i=m-N_h}^m E_{\xi_{k:i}} \left\{ [\mathbf{y}_i - \mathbf{G}_k(\xi_{k:i})\hat{\mu}_{k:i}^\ell] \right. \\ &\quad \left. \times [\mathbf{y}_i - \mathbf{G}_k(\xi_{k:i})\hat{\mu}_{k:i}^\ell]^H | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \quad (24) \end{aligned}$$

due to the fact that the backward update of  $\hat{\mu}_{k:i}$  only covers the region  $[m, \dots, m-N_h]$ .

### C. Forward-Recursion Channel Updates

To further reduce the complexity, the extent of the backward update can be limited to a value between 0 and  $N_h$  in (21) and (23). For the most simple case where no backward update is employed, the estimates become

$$\begin{aligned} \hat{\mu}_{k:m}^\ell &= \mathbf{F}_k^H \hat{\mu}_{k:m-1}^\ell + \Sigma_{m|m-1} \left( \tilde{\mathbf{C}}_{k:m}^{-1} + \Sigma_{m|m-1} \right)^{-1} \\ &\quad \times \left( \tilde{\mu}_{k:m} - \mathbf{F}_k^H \hat{\mu}_{k:m-1}^\ell \right), \quad (25) \end{aligned}$$

$$\hat{\mathbf{C}}_{k:m} = \left(1 - \frac{1}{m}\right) \hat{\mathbf{C}}_{k:m-1} + \frac{1}{m} \Delta \tilde{\mathbf{C}}_{k:m} \quad (26)$$

where

$$\begin{aligned} \Delta \tilde{\mathbf{C}}_m &\triangleq E_{\xi_{k:m}} \left\{ [\mathbf{y}_m - \mathbf{G}_k(\xi_{k:m})\hat{\mu}_{k:m}^\ell] \right. \\ &\quad \left. \times [\mathbf{y}_m - \mathbf{G}_k(\xi_{k:m})\hat{\mu}_{k:m}^\ell]^H | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \end{aligned}$$

and  $\Sigma_{m|m-1} \triangleq \mathbf{B}_k \mathbf{B}_k^H + \mathbf{F}_k^H \mathbf{P}_{m-1}^\ell \mathbf{F}_k$ . These expressions provide more insight into the blind recursive estimator.

Observe that  $\tilde{\mu}_{k:m} = \tilde{\mathbf{C}}_{k:m}^{-1} \tilde{\mathbf{S}}_{k:m}^H \mathbf{y}_m$  is obtained from minimizing

$$\begin{aligned} E_{\psi_{k:m}} \left\{ [\mathbf{y}_m - \mathbf{G}_k(\xi_{k:m})\mathbf{h}_{k:m}]^H \mathbf{C}_k^{-1} \right. \\ \left. \times [\mathbf{y}_m - \mathbf{G}_k(\xi_{k:m})\mathbf{h}_{k:m}] | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\ell-1} \right\} \quad (27) \end{aligned}$$

which is essentially the *stochastic ML channel estimate* of  $\mu_{k:m}$  based on the current observation  $\mathbf{y}_m$  only. The channel estimate  $\hat{\mu}_{k:m}^\ell$  in (25) is equal to the prediction,  $\mathbf{F}_k^H \hat{\mu}_{k:m-1}^\ell$ , given by the channel model plus a correction term proportional to the difference between the stochastic ML estimate and the prediction. Furthermore, the Kalman gain of this estimator is  $\mathbf{K}_{k:m} \triangleq \Sigma_{m|m-1} (\tilde{\mathbf{C}}_{k:m}^{-1} + \Sigma_{m|m-1})^{-1}$ . It is straightforward to show that the eigenvalues of  $\mathbf{K}_{k:m}$ , denoted by  $\Lambda(\mathbf{K}_{k:m})$ , satisfy  $\mathbf{0} \leq \Lambda(\mathbf{K}_m) \leq \mathbf{I}$ . If there is no noise present in the system, then the eigenvalues of  $\tilde{\mathbf{C}}_{k:m}^{-1}$  are close to zero, and thus  $\Lambda(\mathbf{K}_{k:m}) \simeq \mathbf{I}$ .

Under this circumstance, we have  $\hat{\mu}_{k:m}^\ell = \tilde{\mu}_{k:m}$ . This implies that the stochastic ML estimate, based on  $\mathbf{y}_m$  only, suffices to provide accurate channel estimates. On the other hand, if the noise strength is extremely high such that  $\Lambda(\mathbf{K}_{k:m}) \simeq \mathbf{0}$ , then it becomes  $\hat{\mu}_{k:m}^\ell = \mathbf{E}_k^H \hat{\mu}_{k:m-1}^\ell$  which implies that the estimate obtained from the channel model is more reliable than the statistical estimate  $\tilde{\mu}_{k:m}$ , under this circumstance.

## V. NONCOHERENT DETECTION

The EM algorithm only guarantees convergence to a local maximum [19]. Assume the initial values for the EM iterations fall within a neighborhood of the true channel parameters, then the EM algorithm can effectively track the channel parameters, resulting in reliable channel and symbol information. However, when the channel gain attenuates close to zero, the locally maximal points become very close to each other, (e.g. the same channel gain with an in-phase state and a 180-degree out-of-phase state for BPSK modulation). Under these channel conditions, the EM algorithm becomes vulnerable to the channel noise and may lock onto a false state when the channel regains strength. This false-locking phenomenon is widely known as the phase ambiguity problem for blind channel estimation.

The phase ambiguity is relative steady and may only change its status in deep fading and strong noise. Therefore, neglecting estimation errors, the best achievable EM channel estimate within a processing block can be modeled as  $\hat{\mathbf{h}}_{k:m} = [h_{k0}(m)e^{-j\phi_{k0}}, \dots, h_{k(L_k-1)}(m)e^{-j\phi_{k(L_k-1)}}]^T$ , where  $h_{kl}(m)$  is the exact channel parameter of path  $l$  and  $-\phi_{kl}$  is the corresponding phase error over the processing block. The phase error  $\phi_{kl}$  is not uniformly distributed over  $[-\pi, \pi)$ . It depends on the modulation. For MPSK, it takes values on the discrete set  $\Phi \triangleq \{\phi^i = \frac{2i\pi}{M} : i = 0, \dots, M_s - 1\}$ , where  $M_s$  is the constellation size of MPSK. For more general discussion on the relationship between phase ambiguities and modulation schemes, one can refer to [29].

Common methods for combating phase ambiguity in detection include differential encoding or asymmetric modulation [29]. In this section, based on the noncoherent BCJR algorithm, we develop for MPSK a method to compute the correct APP of  $\xi_{k:m}$  by exploiting the phase error characteristics of the EM channel estimates. Later in Section VI-C, a phase correction scheme will be introduced to recover the channel phases of the EM estimates, incorporating the noncoherent APPs provided herein. This phase information plays a crucial role in interference cancellation, without which MAI can not be suppressed by cancellation even if all interfering users' channel gains and transmitted symbols are accurately acquired by noncoherent methods.

### A. Noncoherent Detection Over Multipath Fading Channels

In contrast to [30] which considers noncoherent detection over AWGN channels, we perform herein noncoherent detection over multipath fading channels with partial channel state information (CSI), namely only with the correct estimates of channel gains. Hence, the noncoherent BCJR originally developed in [30] for AWGN channels under continuous phase errors is modified to incorporate CSI and discrete

random phase errors. We assume that CSI can be estimated up to  $\mathbf{h}_k = [\hat{\mu}_{k0}(m)e^{j\phi_{k0}}, \dots, \hat{\mu}_{k(L_k-1)}(m)e^{j\phi_{k(L_k-1)}}]$ , where  $\hat{\mu}_{k:m}$  is obtained with the EM algorithm and  $-\underline{\phi}_k \triangleq [-\phi_{k0}, \dots, -\phi_{k(L_k-1)}]^T$  is the inherent phase error vector for the EM estimates. As was done in [30], we also assume that the phase errors are fixed within a processing block and have memory to some extent. Therefore, in addition to the channel memory resulting from arrival delays, a phase memory of length  $p_k \geq 0$  is also introduced to characterize the phase errors. We define the extended symbol state  $\zeta_{k:m} \triangleq \{\varphi_{k:m}, v_{k:m-p_k}\} = \{b_{k:m-1}, \dots, b_{k:m-q_k}\}$ , where  $\varphi_{k:m} \triangleq \{b_{k:m-1}, \dots, b_{k:m-p_k}\}$  is attributed to the phase memory, and the symbol state  $v_{k:m-p_k} \triangleq \{b_{k:m-p_k-1}, \dots, b_{k:m-q_k}\}$  is due to the channel memory. The total length of the extended channel memory is  $q_k = p_k + L_k^p - 1$ , where  $L_k^p - 1$  is the length of the original channel memory. A diagram is shown in Fig. 2 to illustrate the relationship between  $\varphi_{k:m}$  and  $v_{k:m-p_k}$ .

We next evaluate the APP of  $b_{k:m}$  without correct phase information. By Bayes' rule, we have

$$\begin{aligned} P(b_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}) &= P(\mathbf{Y}_M | b_{k:m}; \hat{\Theta}_{k:M}) \\ &\quad \times P(\mathbf{Y}_M; \hat{\Theta}_{k:M}) P(b_{k:m}) \\ &\propto P(\mathbf{Y}_M | b_{k:m}; \hat{\Theta}_{k:M}). \end{aligned} \quad (28)$$

The memory length constraint imposed on the phase error allows us to exploit the conditional independence structure revealed in  $P(\mathbf{Y}_M | b_{k:m}; \hat{\Theta}_{k:M})$ . Let  $\mathbf{y}_{m_1}^{m_2}$  denote  $\{\mathbf{y}_{m_2}, \dots, \mathbf{y}_{m_1}\}$ . Due to the memory length constraint, the observations after time  $m$ , i.e.  $\mathbf{y}_{m+1}^M$ , are independent of the observations back in time beyond the extent of phase memory, namely  $\mathbf{y}_1^{m-p_k-1}$ . As a result,

$$\begin{aligned} P(\mathbf{Y}_M | b_{k:m}; \hat{\Theta}_{k:M}) &= \sum_{\zeta_{k:m}} P(\mathbf{Y}_M, \zeta_{k:m} | b_{k:m}; \hat{\Theta}_{k:M}) \\ &= \sum_{\zeta_{k:m}} \left\{ P(\mathbf{y}_{m+1}^M | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \right. \\ &\quad \times P(\mathbf{y}_1^{m-p_k-1} | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \\ &\quad \left. \times P(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) P(\zeta_{k:m}) \right\}. \end{aligned} \quad (29)$$

To obtain the APP of  $b_{k:m}$ , the above individual probabilities must be computed in advance.

In the absence of  $\underline{\phi}_k \triangleq [\phi_{k0}, \dots, \phi_{k(L_k-1)}]^T$ , the likelihood of  $\{b_{k:m}, \zeta_{k:m}\}$  is

$$\begin{aligned} \gamma(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) &\triangleq P(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \\ &= \sum_{\underline{\phi}_k} P(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}, \underline{\phi}_k; \hat{\Theta}_{k:M}) \\ &\quad \times P(\underline{\phi}_k) \end{aligned} \quad (30)$$



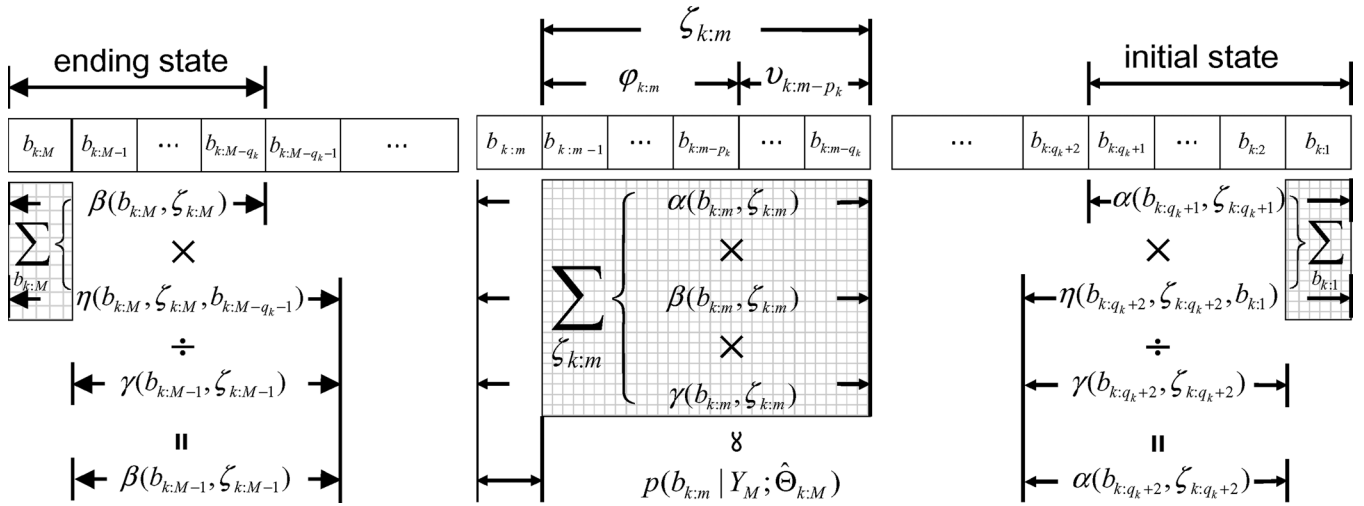


Fig. 2. Data flow diagram for the noncoherent BCJR algorithm. The evaluation for  $\gamma$  and  $\eta$  as well as the forward recursion of  $\alpha$  proceeds from time  $q_k+2$  and ends at time  $M$ . The backward recursion of  $\beta$  starts from time  $M-1$  back to time  $q_k+2$ . The effective posterior probability of  $b_{k:m}$  is given for  $m = q_k+2, \dots, M - q_k - 1$ . The initial and ending states are fixed and given.

where the summation is taken w.r.t. all possible outcomes of  $\underline{\phi}_k$ , and

$$\begin{aligned}
 P\left(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}, \underline{\phi}_k; \hat{\Theta}_{k:M}\right) &= \frac{1}{[\pi^N \det(\mathbf{C})]^{p_k+1}} \\
 &\times \exp\left\{-\sum_{i=0}^{p_k} \left[\mathbf{y}_{m-i} - \mathbf{G}_k(\xi_{k:m-i}) \hat{\mathbf{H}}_{k:m} e^{j\phi_k}\right]^H \mathbf{C}_k^{-1} \right. \\
 &\quad \left. \times \left[\mathbf{y}_{m-i} - \mathbf{G}_k(\xi_{k:m-i}) \hat{\mathbf{H}}_{k:m} e^{j\phi_k}\right]\right\}, \quad (31)
 \end{aligned}$$

$\hat{\mathbf{H}}_{k:m} \triangleq \text{diag}(\hat{\mathbf{h}}_{k:m})$ . Marginalizing  $\underline{\phi}_k$  out of  $P(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}, \underline{\phi}_k; \hat{\Theta}_{k:M})$  yields  $\gamma(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})$ .

As for  $P(\mathbf{y}_1^{m-p_k-1} | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})$ , it can be expressed in a forward recursion form

$$\begin{aligned}
 \alpha(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) &\triangleq P\left(\mathbf{y}_1^{m-p_k-1} | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right) \\
 &= \sum_{b_{k:m-q_k-1}} P\left(\mathbf{y}_1^{m-p_k-1} | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \right. \\
 &\quad \left. \hat{\Theta}_{k:M}\right) P(b_{k:m-q_k-1}) \quad (32)
 \end{aligned}$$

as in [30], where

$$\begin{aligned}
 P\left(\mathbf{y}_1^{m-p_k-1} | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right) &= P\left(\mathbf{y}_1^{m-p_k-2} | \mathbf{y}_{m-p_k-1}^{m-1}, b_{k:m-1}, \zeta_{k:m-1}; \hat{\Theta}_{k:M}\right) \\
 &\quad \times P\left(\mathbf{y}_{m-p_k-1} | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right) \quad (33)
 \end{aligned}$$

due to the fact that  $\{\zeta_{k:m}, b_{k:m-q_k-1}\} \equiv \{b_{k:m-1}, \zeta_{k:m-1}\}$  and that  $\mathbf{y}_m$  and  $b_{k:m}$  are independent of  $\mathbf{y}_1^{m-p_k-2}$ , according to the

memory length assumption. In addition, it follows that

$$\begin{aligned}
 P\left(\mathbf{y}_{m-p_k-1} | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right) &= \frac{P\left(\mathbf{y}_{m-p_k-1}^m | b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right)}{P\left(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right)} \\
 &\triangleq \frac{\eta\left(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right)}{\gamma\left(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right)}. \quad (34)
 \end{aligned}$$

Similar to the derivations in (30), in the absence of  $\underline{\phi}_k$ ,

$$\begin{aligned}
 \eta\left(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right) &\triangleq \sum_{\underline{\phi}_k} \\
 &\quad \times P\left(\mathbf{y}_{m-p_k-1}^m | b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}, \underline{\phi}_k; \hat{\Theta}_{k:M}\right) P(\underline{\phi}_k). \quad (35)
 \end{aligned}$$

Therefore, the forward recursion of  $\alpha$  is given by

$$\begin{aligned}
 \alpha(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) &= \sum_{b_{k:m-q_k-1}} \alpha\left(b_{k:m-1}, \zeta_{k:m-1}; \hat{\Theta}_{k:M}\right) \\
 &\quad \times \frac{\eta\left(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}\right)}{\gamma\left(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right)} P\left(b_{k:m-q_k-1}\right). \quad (36)
 \end{aligned}$$

Finally,  $P(\mathbf{y}_{m+1}^M | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})$  can also be expressed in a backward recursion form

$$\begin{aligned}
 \beta(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) &\triangleq P\left(\mathbf{y}_{m+1}^M | \mathbf{y}_{m-p_k}^m, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right) \\
 &= \sum_{b_{k:m+1}} P\left(\mathbf{y}_{m+1}^M | \mathbf{y}_{m-p_k}^m, b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right) \\
 &\quad \times P(b_{k:m+1}) \quad (37)
 \end{aligned}$$

as in [30], where

$$\begin{aligned} & P\left(\mathbf{y}_{m+1}^M | \mathbf{y}_{m-p_k}^m, b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right) \\ &= P\left(\mathbf{y}_{m+2}^M | \mathbf{y}_{m-p_k+1}^{m+1}, b_{k:m+1}, \zeta_{k:m+1}; \hat{\Theta}_{k:M}\right) \\ & \quad \times P\left(\mathbf{y}_{m+1}^m | \mathbf{y}_{m-p_k}^m, b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right) \end{aligned} \quad (38)$$

due to the fact that  $\{b_{k:m}, \zeta_{k:m}\} \equiv \{\zeta_{k:m+1}, b_{k:m-q_k}\}$ , and that  $\mathbf{y}_{m+2}^M$  is independent of  $\mathbf{y}_{m-p_k}^m$  and  $b_{k:m-q_k}$ , according to the memory length assumption. In addition, we also have

$$\begin{aligned} & P\left(\mathbf{y}_{m+1}^m | \mathbf{y}_{m-p_k}^m, b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right) \\ &= \frac{P\left(\mathbf{y}_{m-p_k}^{m+1} | b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right)}{P\left(\mathbf{y}_{m-p_k}^m | b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}\right)} \\ &\triangleq \frac{\eta(b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})}{\gamma(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})}. \end{aligned} \quad (39)$$

Thus, the backward recursion of  $\beta$  is given by

$$\begin{aligned} \beta(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) &= \sum_{b_{k:m+1}} \beta(b_{k:m+1}, \zeta_{k:m+1}; \hat{\Theta}_{k:M}) \\ & \quad \times \frac{\eta(b_{k:m+1}, b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})}{\gamma(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})} P(b_{k:m+1}). \end{aligned} \quad (40)$$

Now, substituting (30), (36) and (40) back into (29) yields

$$\begin{aligned} P(\mathbf{Y}_M | b_{k:m}; \hat{\Theta}_{k:M}) &= \sum_{\zeta_{k:m}} \alpha(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \\ & \quad \times \gamma(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \beta(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) P(\zeta_{k:m}). \end{aligned} \quad (41)$$

The signal processing procedure for calculating the noncoherent posterior probability is illustrated in Fig. 2. To initiate the forward and backward recursions of (36) and (40), respectively, both the initial state  $\{b_{k:q_k+1}, \zeta_{k:q_k+1}\}$  and the ending state  $\{b_{k:M}, \zeta_{k:M}\}$  must be given *a priori*. In practice, this information can be inserted in packet headers. Other than the complexity imposed by the extended channel memory length, the complexity of this algorithm is dominated by the summations involved in (30) and (35), which seem to have no closed-form solutions except for the flat fading case that will be discussed in the next section. Taking into account the complexity of (31), the overall complexity of this algorithm is  $O(M_s^{L_k} L_k^2)$ .

### B. Noncoherent Detection Over Flat Fading Channels

For flat fading channels,  $L_k = 1$  and  $L_k^p = 2$ . Thus (31) degenerates to

$$\begin{aligned} & P\left(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}, \phi_k; \hat{\Theta}_{k:M}\right) \\ &= \frac{1}{[\pi^N \det(\mathbf{C}_k)]^{p_k+1}} \\ & \quad \cdot \exp\left\{-\sum_{i=0}^{p_k} \left(\mathbf{y}_{m-i} - \mathbf{G}_k(\xi_k^{m-i}) \hat{h}_k e^{j\phi_k}\right)^H \mathbf{C}_k^{-1} \right. \\ & \quad \left. \times \left(\mathbf{y}_{m-i} - \mathbf{G}_k(\xi_k^{m-i}) \hat{h}_k e^{j\phi_k}\right)\right\} \end{aligned} \quad (42)$$

where  $\mathbf{G}_k = \mathbf{g}_k$  is now a vector. Marginalizing out the phase error  $\phi_k$  gives

$$\begin{aligned} & \gamma(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \\ &\triangleq \sum_{\phi_k} P\left(\mathbf{y}_{m-p_k}^m | b_{k:m}, \zeta_{k:m}, \hat{\Theta}_{k:M}, \phi_k\right) \\ &= \frac{\exp\left\{D_k(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})\right\}}{[\pi^N \det(\mathbf{C}_k)]^{p_k+1}} \cdot \frac{1}{M_s} \\ & \quad \times \sum_{i=0}^{M_s} \exp\left\{2 \left|F_k(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})\right| \cos(\phi_k^i + \omega_k)\right\} \end{aligned} \quad (43)$$

where

$$\begin{aligned} D_k(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) &\triangleq -\sum_{i=0}^{p_k} [\mathbf{y}_{m-i}^H \mathbf{C}_k^{-1} \mathbf{y}_{m-i} \\ & \quad + \hat{h}_k^H \mathbf{G}_k^H(\xi_k^{m-i}) \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_k^{m-i}) \hat{h}_k], \end{aligned}$$

$F_k(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \triangleq \sum_{i=0}^{p_k} \mathbf{y}_{m-i}^H \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_k^{m-i}) \hat{h}_k$  and  $\omega_k$  is the phase angle of  $F_k(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})$ . The above expression can be further simplified since the signal constellation of MPSK is formed with conjugate pairs on a circle, with 180° phase difference in each pair. Therefore, (43) becomes

$$\begin{aligned} & \gamma(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M}) \\ &= \frac{2 \exp\left\{D_k(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})\right\}}{M_s [\pi^N \det(\mathbf{C}_k)]^{p_k+1}} \\ & \quad \times \sum_{i=0}^{\frac{M_s}{2}-1} \cosh\left\{2 \left|F_k(b_{k:m}, \zeta_{k:m}; \hat{\Theta}_{k:M})\right| \right. \\ & \quad \left. \times \cos\left(\frac{4\pi i}{M_s} + \omega_k\right)\right\}. \end{aligned} \quad (44)$$

Similarly, for  $\eta$  we have

$$\begin{aligned} & \eta(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}) \\ &= \frac{2 \exp\left\{D_k(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M})\right\}}{M_s [\pi^N \det(\mathbf{C}_k)]^{p_k+2}} \\ & \quad \cdot \sum_{i=0}^{\frac{M_s}{2}-1} \cosh\left\{2 \left|F_k(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M})\right| \right. \\ & \quad \left. \times \cos\left(\frac{4\pi i}{M_s} + \omega_k\right)\right\} \end{aligned} \quad (45)$$

where

$$\begin{aligned} & D_k(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}) \\ &\triangleq -\sum_{i=0}^{p_k+1} [\mathbf{y}_{m-i}^H \mathbf{C}_k^{-1} \mathbf{y}_{m-i} \\ & \quad + \hat{h}_k^H \mathbf{G}_k^H(\xi_k^{m-i}) \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_k^{m-i}) \hat{h}_k] \end{aligned}$$

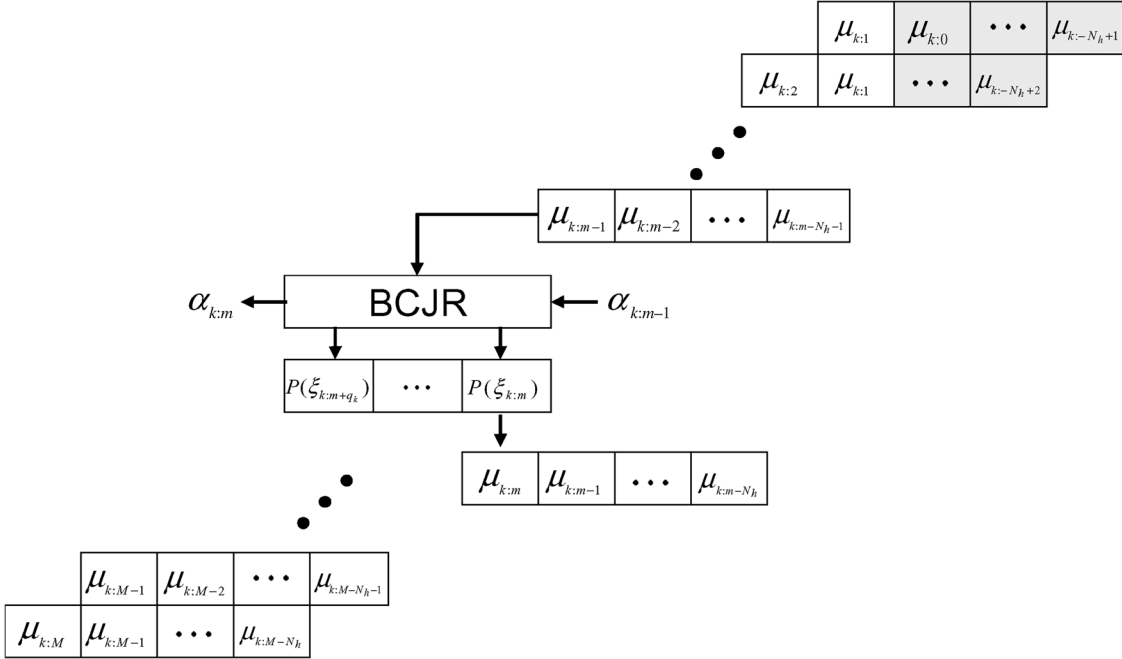


Fig. 3. Signal processing procedure for exploring unknown channel state information.

and

$$F_k(b_{k:m}, \zeta_{k:m}, b_{k:m-q_k-1}; \hat{\Theta}_{k:M}) \triangleq \sum_{i=0}^{p_k+1} \mathbf{y}_{m-i}^H \mathbf{C}_k^{-1} \mathbf{G}_k(\xi_k^{m-i}) \hat{h}_k.$$

We note that the APP of  $b_{k:m}$  is still given by (41).

## VI. SIGNAL PROCESSING PROCEDURES FOR JOINT CHANNEL ESTIMATION AND SYMBOL DETECTION

We now have the necessary tools ready for joint channel estimation and symbol detection. The remaining problem is how to apply the above results in a systematic way to track a unknown time-varying channel. This is more involved than it appears to be. To help interpret the entire signal processing procedures, the whole process is partitioned into three consecutive steps which are referred to, respectively, as the channel exploration, channel refinement and phase adjustment. We first introduce the channel exploration step to obtain a rough channel estimate for a time-varying channel without the prior knowledge of transmitted data.

### A. Channel Exploration

The channel exploration can start with or without initial CSI. With initial CSI, the algorithm will converge faster. The availability of the initial CSI does not affect the final estimation results. Without loss of generality, we assume the initial CSI state,  $\underline{\mu}_{k:0} \triangleq \{\mu_{k:0}, \dots, \mu_{k:-N_h+1}\}$  as shown in the shaded region in Fig. 3, is obtained either with a short training sequence or by random assignment. On the other hand, the initial covariance is set to  $\mathbf{C}_{k:0} \triangleq \mathbf{I}$ .

In the absence of both CSI and transmitted data, CSI has to be first explored step by step from time 1 to time  $M$ .

For every new time step  $m$ , the initial estimate is given by  $\hat{\mu}_{k:m|m-1} \triangleq \mathbf{F}_k^H \hat{\underline{\mu}}_{k:m-1}^{\mathcal{L}}$  using the dynamic model (11). In addition, for each time step, the EM procedure is iterated for a number of times, denoted by  $\mathcal{L}$ , before reaching a steady-state channel estimate.<sup>5</sup> Thus, the initial parameter set is  $\hat{\Theta}_{k:m}^0 \triangleq \{\hat{\mu}_{k:m|m-1}, \hat{\Theta}_{k:m-1}^{\mathcal{L}}\}$ , and the estimates at the  $\ell$ th iteration are  $\hat{\Theta}_{k:m}^{\ell} \triangleq \{\hat{\mu}_{k:m}^{\ell}, \hat{\mathbf{C}}_{k:m}^{\ell}, \hat{\mathcal{U}}_{k:m-1}^{\ell}\}$ .

To initiate the channel tracking, we set  $\hat{\Theta}_{k:1}^0 \triangleq \{\hat{\mu}_{k:1|0}, \mathbf{C}_{k:0}\}$  and use a sliding-window BCJR algorithm to evaluate  $P(\xi_{k:1} | \mathbf{y}_1^{d_{Lag}+1}; \hat{\Theta}_{k:1}^0)$ . Due to ignorance of future CSI, the channel parameter used for the sliding-window BCJR algorithm is set to  $\hat{\mu}_{k:1|0}^0$ . Similarly, the observations used for the BCJR are  $\mathbf{y}_1, \dots, \mathbf{y}_{1+d_{Lag}}$ , where  $d_{Lag}$  is the smoothing lag set to isolate the future observations from the current processing block. The size of  $d_{Lag}$  is a design parameter. A small value of it will affect the accuracy of  $P(\xi_{k:1} | \mathbf{y}_1^{d_{Lag}+1}; \hat{\Theta}_{k:1}^0)$ , whereas a large value will make the use of  $\hat{\mu}_{k:1|0}^0$  less realistic for the entire processing block. A reasonable value should be less than the coherence time of a fading channel. In the simulation studies,  $d_{Lag}$  is set equal to the length of channel memory,  $q_k$ .

After obtaining  $P(\xi_{k:1} | \mathbf{y}_1^{d_{Lag}+1}; \hat{\Theta}_{k:1}^0)$ , we substitute it back into (14) and (15), which results in the new parameters  $\hat{\mathbf{S}}_{k:1}$  and  $\hat{\mathbf{C}}_{k:1}$ . With these two parameters as well as the channel update equation (21), we immediately obtain  $\hat{\underline{\mu}}_{k:1}^1$ , where the initial matrix  $\mathbf{P}_{k:0}^0 \propto \mathbf{I}$ . Similarly, given  $\hat{\underline{\mu}}_{k:1}^1$ , the estimate  $\hat{\mathbf{C}}_{k:1}^1$  can also be obtained with (23) and (24).

The aforementioned procedure can be repeated, by the EM principle, until converging to a steady state. The corresponding steady-state estimates are denoted as  $\Theta_{k:1}^{\mathcal{L}} = \{\hat{\underline{\mu}}_{k:1}^{\mathcal{L}}, \hat{\mathbf{C}}_{k:1}^{\mathcal{L}}\}$ .

<sup>5</sup>Our simulation studies show that one or two iterations are sufficient for channel exploration. More iterations do not help due to the limited observations available for channel exploration.

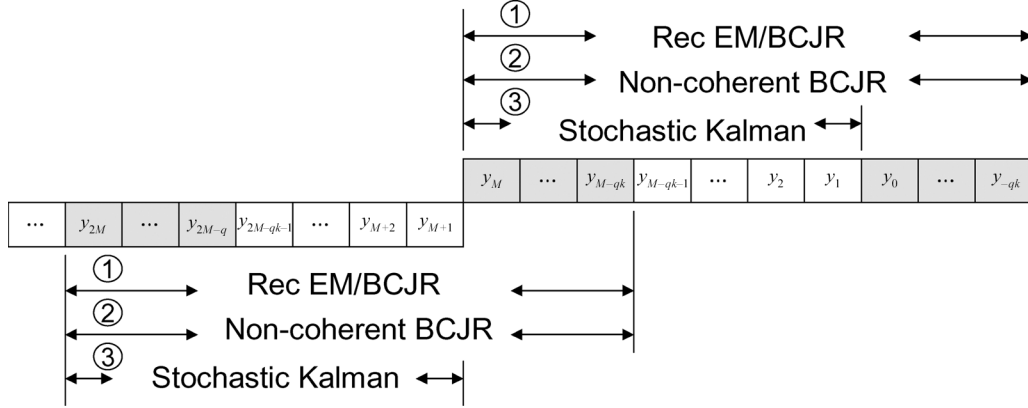


Fig. 4. Data blocks for the signal processing of joint channel estimation and MAP symbol detection. The first block is for channel exploration and channel refinement, the second for evaluating APP using the noncoherent BCJR and the third for channel adjustment with the stochastic Kalman estimator (21).

Continuing the same procedure for each processing window  $\{\mathbf{y}_m, \dots, \mathbf{y}_{m+d_{Lag}}\}$ ,  $m = 2, \dots, M$ , leads to the steady-state estimates  $\{\hat{\mu}_{k:m}^{\mathcal{L}}, \hat{\mathbf{C}}_{k:m}^{\mathcal{L}}\}$ , and hence  $\hat{\Theta}_{k:M}^{\mathcal{L}} = \{\hat{\mathbf{U}}_{k:M}^{\mathcal{L}}, \hat{\mathbf{C}}_{k:M}^{\mathcal{L}}\}$  of the channel exploration phase. The signal processing flow for channel exploration is illustrated in Fig. 3.

It is noted that upon acquisition of the update  $\{\hat{\mu}_{k:m}^{\mathcal{L}}, \hat{\mathbf{C}}_{k:m}^{\mathcal{L}}\}$ , the sliding-window BCJR algorithm is executed one more time to obtain a new  $P(b_{k:m} | \mathbf{y}_m^{m+d_{Lag}}; \hat{\Theta}_{k:m}^{\mathcal{L}})$  with the new set  $\hat{\Theta}_{k:m}^{\mathcal{L}}$ . This will lead to a new K-L measure  $Q_m(\Theta_{k:m} | \hat{\Theta}_{k:m}^{\mathcal{L}})$ , and, in turn, another new update  $\{\hat{\mu}_{k:m}^{\mathcal{L}+1}, \hat{\mathbf{C}}_{k:m}^{\mathcal{L}+1}\}$ . The likelihood of  $\hat{\Theta}_{k:m}^{\mathcal{L}}$  is non-decreasing by the recursive EM framework [10].

### B. Channel Refinement

In the channel exploration phase, due to the ignorance of the future channel parameters, a sliding-window BCJR algorithm of size  $d_{Lag} + 1$  is used for the evaluation of  $P(\xi_{k:m} | \mathbf{y}_m^{m+d_{Lag}}; \hat{\Theta}_{k:m}^{\mathcal{L}-1})$ , which may make the APP of  $\xi_{k:m}$  less reliable, and, consequently, yield less accurate estimates of  $\{\hat{\mu}_{k:m}^{\mathcal{L}}, \hat{\mathbf{C}}_{k:m}^{\mathcal{L}}\}$ . To improve the performance of channel tracking, the entire process of the EM iteration which involves BCJR and the stochastic recursive estimators (21) and (23), can be redone on a block basis of size  $M$  as shown in Fig. 4, using  $\hat{\Theta}_{k:M}^{\mathcal{L}}$  obtained from the channel exploration phase as the initial parameter set. To help track the time-varying channel parameters, adjacent blocks are overlapped to some extent. The parameters in the overlapped region at the end of a block serve as the initial parameters for channel exploration of the next processing block.

Within each block, starting with  $\hat{\Theta}_{k:M}^{\mathcal{L}}$ , a more reliable  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\mathcal{L}})$  can be obtained, owing to the full observation  $\mathbf{Y}_M$  of the current block. In addition, with the new updated APP, the recursive estimators of (21) and (23) can be executed once again from time 0 to time  $M$  to obtain a refined estimate of  $\hat{\Theta}_{k:M}^{\mathcal{L}+1}$ . By the EM principle,  $\mathcal{L}\mathcal{L}\mathcal{K}(\hat{\Theta}_{k:M}^{\mathcal{L}+1}) \geq \mathcal{L}\mathcal{L}\mathcal{K}(\hat{\Theta}_{k:M}^{\mathcal{L}})$ . Similarly, this EM procedure can be repeated a number of times, with the final steady-state estimate denoted by  $\hat{\Theta}_{k:M}$ . The combination of the channel exploration phase and the channel refinement phase is shown as the first stage of signal processing procedures in Fig. 4. The mechanism for channel phase adjustment is presented in the next section.

### C. Phase Adjustment

As pointed out earlier, the EM algorithm only guarantees convergence to a local maximum, which may lead to incorrect APPs and channel estimates with phase errors. For the purpose of symbol detection only, a corrected APP can be obtained with a modified noncoherent BCJR algorithm presented in Section V, using the EM channel estimates. However, when one intends to improve the performance of MUD via interference cancellation, reliable channel phase information must be obtained for all users. We present in this section a method to adjust the EM channel estimates, which incorporates the noncoherent APP, (41), introduced in Section V.

We observe that the reliability of the EM channel estimates is closely related to that of the APP of the symbol state,  $\xi_{k:m}$ . For (14) and (15) involved in the estimator (21), the matrices  $[\mathbf{G}_k^H(\xi_{k:m})\mathbf{C}_k^{-1}\mathbf{G}_k(\xi_{k:m})]$  and  $[\mathbf{G}_k^H(\xi_{k:m})\mathbf{C}_k^{-1}\mathbf{y}_m]$  for every possible state of  $\xi_{k:m}$  are weighted and summed by its corresponding  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\mathcal{L}-1})$ . If the transmitted symbol is known *a priori* as in a training mode, then the estimate  $\tilde{\mu}_{k:m}$  in (21) simplifies to

$$\tilde{\mu}_{k:m} = [\mathbf{G}_k^H(\xi_{k:m})\mathbf{C}_k^{-1}\mathbf{G}_k(\xi_{k:m})]^{-1} [\mathbf{G}_k^H(\xi_{k:m})\mathbf{C}_k^{-1}\mathbf{y}_m] \quad (46)$$

which will lead to a regular Kalman estimator. Therefore, if  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M}^{\mathcal{L}-1})$  in (14) and (15) of (21) are robust to phase errors, then the channel estimators are able to track the channel phase coherently. This phase robustness of APP can be achieved with the noncoherent scheme presented in (41), given the EM channel estimates.

To adjust the channel estimate,  $\hat{\mu}_{k:m}$  is first applied in (31), which is required by (30) and (35) to evaluate  $P(\mathbf{Y}_M | b_{k:m}; \hat{\Theta}_{k:M})$  with (41). The resultant  $P(b_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M})$ , which is proportional to  $P(\mathbf{Y}_M | b_{k:m}; \hat{\Theta}_{k:M})$ , is robust to phase errors. Substituting this probability back into (21) gives the adjusted channel estimate. We note that the phase error coupled to the EM estimates only changes under the circumstance of a deep channel fade and strong noise, otherwise it will stay at the same value until the next adverse channel condition occurs. By correcting the phase error block by block, the stochastic recursive estimator is able to track channel parameters without the exact symbol

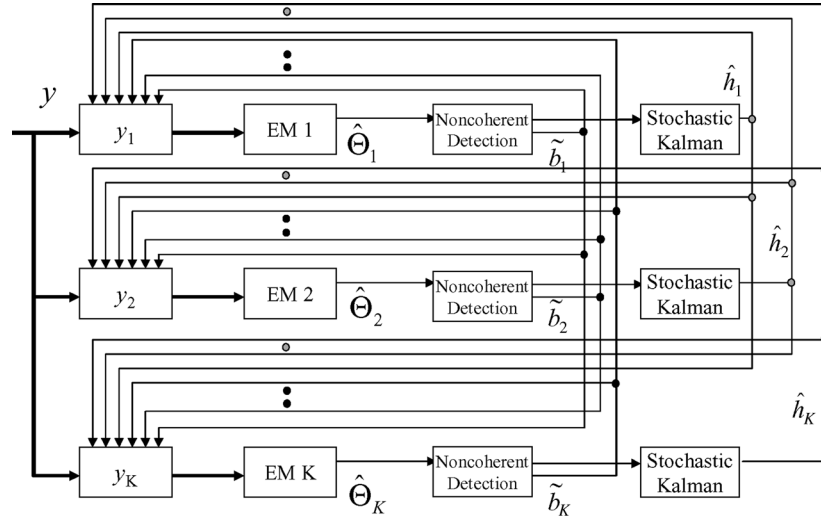


Fig. 5. Signal processing diagram for joint channel estimation and multiuser detection, where the stochastic Kalman refers to the stochastic recursive estimator of (21).

information. A block diagram is presented in Fig. 4 to illustrate the sliding-window mechanism for joint channel estimation, noncoherent detection as well as channel adjustment. The boundary conditions, shown as the gray areas in Fig. 4, must be given *a priori* for the noncoherent BCJR algorithm, which, in practice, can be obtained from packet headers.

## VII. ITERATIVE INTERFERENCE CANCELLATION AND MULTIUSER DETECTION

Soft interference cancellation has been shown in [27] to be an effective low-complexity approach for achieving the single-user detection performance in DS-CDMA systems, given perfect CSI. In addition to reliable CSI, performing soft interference cancellation also requires the soft output of the transmitted symbol of all users. For BPSK, the soft output of each symbol is given by

$$\tilde{b}_{k:m} \triangleq \frac{P(b_{k:m} = 1 | \mathbf{Y}_M; \hat{\Theta}_{k:M}) - P(b_{k:m} = -1 | \mathbf{Y}_M; \hat{\Theta}_{k:M})}{P(b_{k:m} = 1 | \mathbf{Y}_M; \hat{\Theta}_{k:M}) + P(b_{k:m} = -1 | \mathbf{Y}_M; \hat{\Theta}_{k:M})} \quad (47)$$

where  $P(b_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M})$  is obtained by the modified noncoherent BCJR algorithm (41). Given  $\tilde{b}_{k:m}$  of all users, the soft estimate of the MAI,  $\mathbf{i}_k$ , of user  $k$ , in (6) is obtained by replacing the user symbol  $b_{i:m}$ ,  $i \neq k$ , embedded in  $\mathbf{i}_k(m)$  with the soft output  $\tilde{b}_{i:m}$ , as well as substituting  $\hat{\mathbf{h}}_{i:m}$  for  $\mathbf{h}_{i:m}$ . The resultant matrix  $\mathbf{G}_i$  is denoted by  $\mathbf{G}_i(\tilde{\xi}_{i:m})$ , where  $\tilde{\xi}_{i:m} = \{\tilde{b}_{i:m}, \dots, \tilde{b}_{i:m-L_i^p}\}$ , [cf. (6)]. Then, the received signal  $\mathbf{y}_k(m)$  for user  $k$  after interference cancellation is given by

$$\mathbf{y}_k(m) = \mathbf{y}(m) - \sum_{i=0, i \neq k}^{K-1} \mathbf{G}_i(\tilde{\xi}_{i:m}) \hat{\mathbf{h}}_{i:m}. \quad (48)$$

By using  $\mathbf{y}_k$  for the joint symbol detection and channel estimation for user  $k$ , the quality of  $\hat{\Theta}_{k:M}$  and  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M})$  can both be improved due to the suppressed MAI. With more precise estimates of  $\xi_{k:m}$  and  $\mathbf{h}_{k:m}$ , one can obtain better estimate of  $\mathbf{i}_k(m)$ , too. Through iterations, both MAI and the

variance of estimation can be greatly reduced, hence resulting in much better estimation and detection performance than a single-user method.

The entire signal processing flow for joint channel estimation and MAP MUD is presented in Fig. 5. The received signal for one user is first sent to the EM block to compute the channel estimates. The EM block not only involves the stochastic estimators (21), it also employs the *coherent* BCJR algorithm to evaluate  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M})$ . Once the EM algorithm is done with its iterations, the channel estimates  $\hat{\Theta}_{k:M}$  are forwarded to the noncoherent detection block to evaluate a robust version of  $P(\xi_{k:m} | \mathbf{Y}_M; \hat{\Theta}_{k:M})$  using the noncoherent BCJR algorithm (41). The robust APP is further fed to (21) (denoted by stochastic Kalman in the figure) to correct the phase errors coupled on the EM estimates. Notice that no iteration is required between the EM block and the noncoherent BCJR algorithm. The entire procedure for channel re-estimation, symbol re-detection and interference cancellation can be repeated until  $\|\hat{\mathcal{H}}_{k:M}^{\mathcal{L}-1} - \hat{\mathcal{H}}_{k:M}^{\mathcal{L}}\|^2 < \varepsilon$ , for some prespecified tolerance,  $\varepsilon$ . The hard decision only employs at the last stage. For BPSK, it is  $\hat{b}_{k:m} = \text{sgn}[P(b_{k:m} = 1 | \mathbf{Y}_M; \hat{\Theta}_{k:M}) - P(b_{k:m} = -1 | \mathbf{Y}_M; \hat{\Theta}_{k:M})]$ .

## VIII. SIMULATION RESULTS

Computer simulations are conducted to examine the performance of the proposed joint channel estimation and MUD scheme in multipath fading channels. A DS-CDMA system of seven users,  $K = 7$ , is investigated, with the spreading gain  $N = 15$ . Spreading sequences are randomly generated and assigned to users in the system. The variance of the noise for each chip interval is normalized to 2, one for the real part and the other for the imaginary part. The number of channel paths is set to  $L_k = 3$  for each user, with the path delays being generated with a uniform distribution over  $[0, T_s)$ . The delays are assumed pre-acquired and provided to the MUD.

The channel coefficients for each path are assumed Rayleigh and generated following the method in [35]. In this simulation study, we investigate two cases, one with the normalized

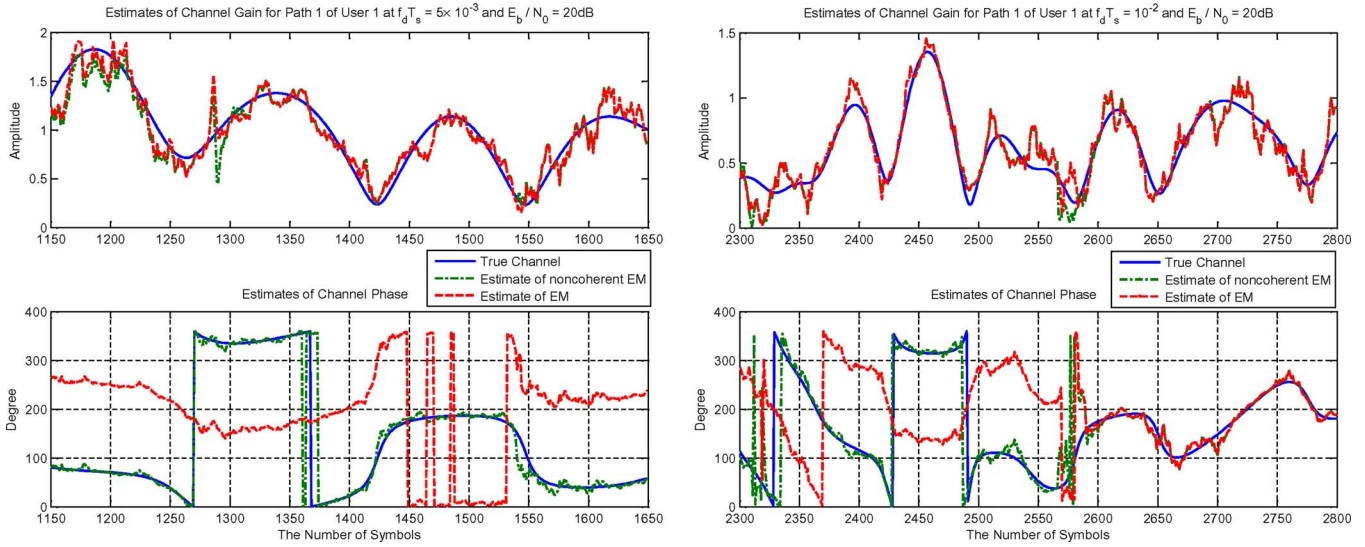


Fig. 6. (left) Simulation results of channel tracking for path 1 of user 1 over a three-path fading channel at  $f_d T_s = 5 \times 10^{-3}$  and (right) path 1 of a three-path fading channel at  $f_d T_s = 10^{-2}$ .

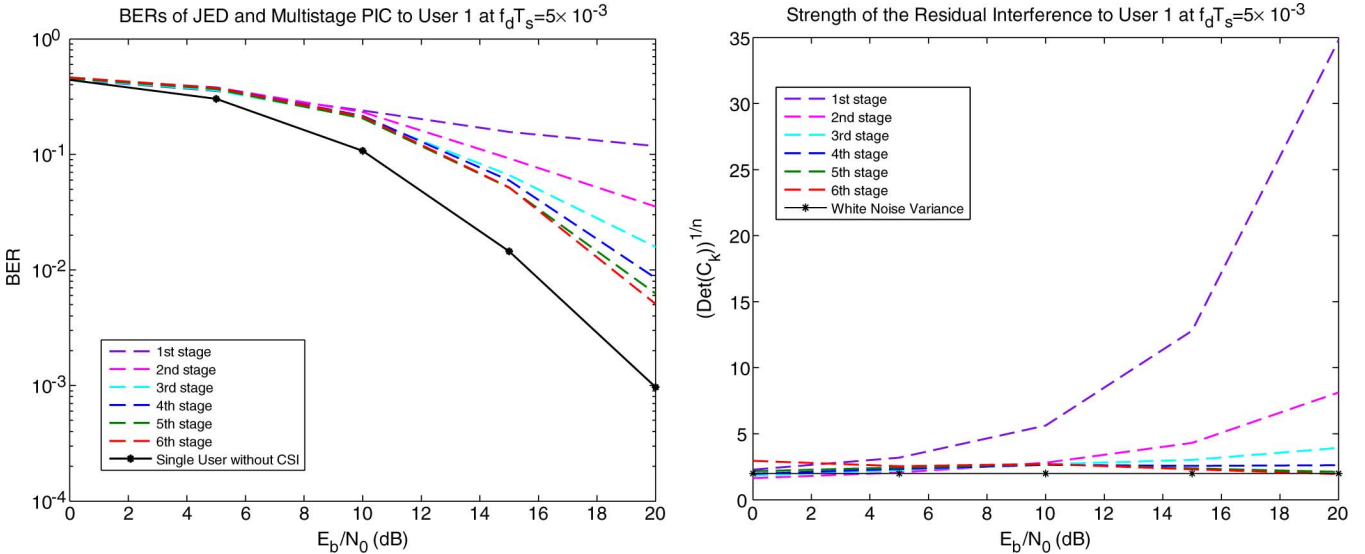


Fig. 7. Effects of JED and multistage PIC on the BER and the residual interference to user 1 of a multiuser system in which  $K = 7$ ,  $N = 15$ ,  $L = 3$ , and  $f_d T_s = 5 \times 10^{-3}$ .

Doppler shift  $f_d T_s = 5 \times 10^{-3}$ , the other with  $f_d T_s = 0.01$ . It is noted that the system under the above simulation setting is overloaded from the perspective of linear multiuser receivers [36] and, as such, MAI can not be effectively suppressed without using interference cancellation.<sup>6</sup>

The proposed joint channel estimation and MUD scheme is performed on a block basis. The sliding-window for channel tracking and phase adjustment contains 128 symbols, and to initialize the noncoherent BCJR algorithm, the first four symbols of each block are used as initial states. The number of iterations for each time step is three, one for channel exploration and the other two for channel refinement. In addition, six stages of

<sup>6</sup>A system is defined to be overloaded in [36] if  $\text{Cov}[\mathbf{i}_k] > N - L_k$ . While this definition is for linear receivers, we observe that the MAP detector derived in this paper generalizes the linear ML detector derived in [36] (the work in [36] ignores ISI). Thus, the performance of the MAP detector before PIC is still constrained by the loading factor defined above.

parallel interference cancellations (PICs) are employed for each user on a block basis.

We first present the performance of channel tracking at  $E_b/N_0 = 20$  dB, using the simplified stochastic recursive estimator (25). The order,  $N_h$ , of the channel model (11) is set to 3 in the simulations.<sup>7</sup> The estimates of channel amplitudes and channel phases for path 1 of user 1 are shown in Fig. 6 for  $f_d T_s = 5 \times 10^{-3}$  and  $f_d T_s = 10^{-2}$ , respectively. Both of the results prior to (referred to as the EM estimates) and post (referred to as the noncoherent EM estimates, due to the noncoherent APPs used for phase correction) the phase adjustment are shown in the figures. As presented in the figures,

<sup>7</sup>A thorough discussion on order selection for training-based Kalman filtering can be found in [37]. For our JED algorithm, we compared the BERs for the single-user system with different values of  $N_h$  (2 to 5), and found almost no performance improvement for  $N_h \geq 3$ . So, we set  $N_h = 3$  so as not to increase the system complexity.

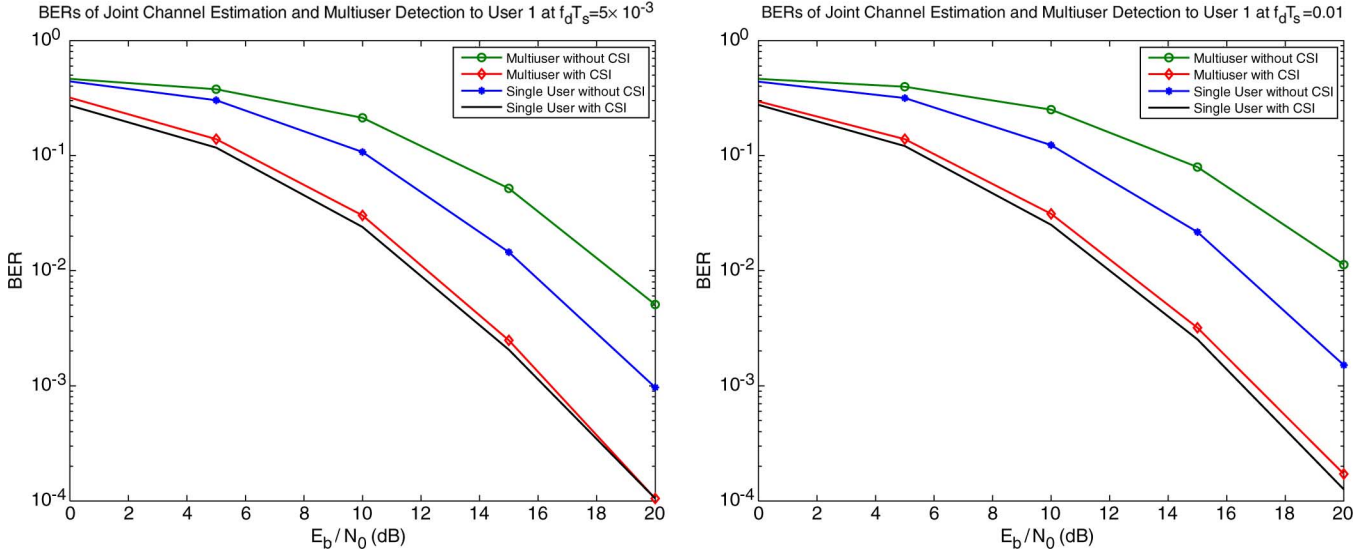


Fig. 8. (left) BERs for  $f_d T_s = 5 \times 10^{-3}$  under different operating conditions; (right) similar results for  $f_d T_s = 10^{-2}$ .

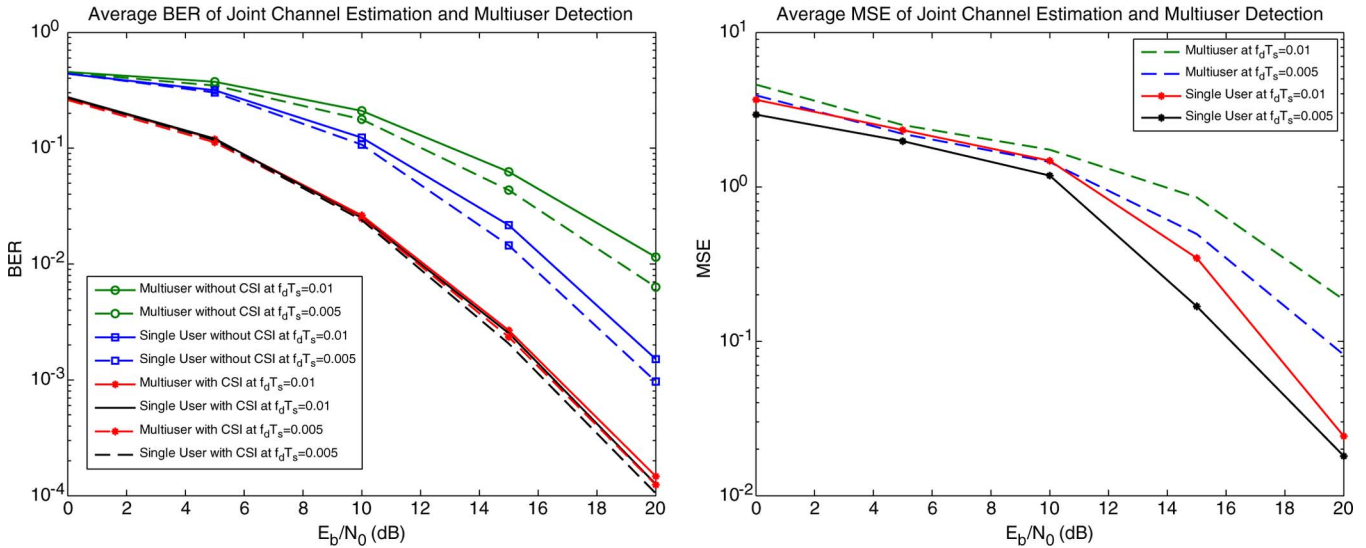


Fig. 9. (left) Average BERs over all users in the system for different fading speeds; (right) corresponding average MSEs.

the channel amplitudes of both the estimates agree with the true channel ones. However, the phase estimates of the EM algorithm often go  $180^\circ$  out of the phase of the true channel. Only the noncoherent EM algorithm provides correct phase estimates. This shows that the phase errors can be recovered for PIC with the proposed phase adjustment scheme.

Fig. 7 demonstrates the effects of JED and multistage PIC on the bit error rate (BER) and the residual interference to user 1 when  $f_d T_s = 5 \times 10^{-3}$ . The effect of multistage PIC on the BER of JED is shown in the left plot, while the right plot presents the strength of the residual MAI plus noise. The strength of the residual interference is measured in  $\sqrt{\det(\mathbf{C}_1)} = (\prod_{i=1}^N \lambda_i)^{1/N}$  with  $\lambda_i$  being the eigenvalues of  $\mathbf{C}_1$ . As shown in the figure, the strength drops quickly close to 2 of the white noise power level after four to five stages of PICs, while the BER curve after 6 stages of PICs still sees a 3 dB loss in  $E_b/N_0$  at  $\text{BER} = 10^{-2}$  in comparison with the single-user performance.

Fig. 8 presents the BERs of JED for user 1 at two different fading speeds. The left plot compares the BER of user 1 after six stages of PICs against the BER of JED of a single-user system, the BER of user 1 given perfect CSI and the BER of a single-user system given CSI when  $f_d T_s = 5 \times 10^{-3}$ . Similar simulation results are provided for  $f_d T_s = 10^{-2}$  in the right plot, which show close agreement with that of  $f_d T_s = 5 \times 10^{-3}$ .

As shown in the figures, with prior CSI, the BER of user 1 comes close to that of the single-user BER, implying that the performance of MAP detection is not sensitive to the colored residual interference when perfect CSI is provided. On the other hand, the BER of JED for the single-user system is around 3 dB inferior to the coherent BER, and has another 3 dB gain against the BER of joint estimation and MUD for user 1. This implies that the performance of channel estimation is rather sensitive to the colored residual interference, despite the fact that the strength of the colored residual interference is comparable to that of the white noise as seen in Fig. 7. For a single-user system

under AWGN, via JED, channel estimation errors introduce to detection a 3 dB loss in signal to noise ratio (SNR). While for a multiuser system, estimation errors seem to be worse due to the colored residual interference, consequently, resulting in additional loss in SNR compared with the single-user one. This can be considered as the additional loss in SNR due to the combination of JED and PIC.

Fig. 9 compares the average BERs and mean squared errors (MSEs) over all users for the cases of  $f_d T_s = 10^{-2}$  and  $f_d T_s = 5 \times 10^{-3}$  after 4 stages of PICs. As expected, with perfect CSI, the performance of MUD is negligibly affected by the fading speeds, while the BERs of JED for  $f_d T_s = 10^{-2}$  are about 1 dB inferior to that of  $f_d T_s = 5 \times 10^{-3}$ . Though not significant, this shows that the performance of JED is also affected by the fading speed to certain extent. Different length processing blocks may be needed for different fading speeds to better tradeoff between the performance, pilot expense and delay constraints.

## IX. CONCLUSION

An iterative structure was proposed for JED and PIC in multipath time-varying channels. Through the proposed stochastic recursive channel estimator and the corresponding phase correction scheme, soft PIC is made possible for joint estimation and MUD over fast fading channels. It was shown by simulation that the strength of MAI can be effectively suppressed within four to five stages of soft PICs and, hence, largely improving the performance of MUD at high SNR.

Despite PIC, a 3 dB loss of SNR was observed in the BERs of JED between a multiuser system and a single-user one. This may result from the sensitivity of channel estimation on the colored residual interference and the coupling effects between channel estimation and symbol detection. Shorter length processing blocks may be adopted to improve the performance. However, a more rigorous theoretical investigation is required to characterize the limiting performance of joint estimation and MUD.

## REFERENCES

- [1] *IEEE Standard For Local and Metropolitan Area Networks Part 16: Air Interface For Fixed and Mobile Broadband Wireless Access System Amendment For Physical and Medium Access Control Layers For Combined Fixed and Mobile Operation in Licensed Bands*, IEEE 802, 2005 [Online]. Available: <http://www.standards.ieee.org/getieee802>
- [2] A. Lampe, "Iterative multiuser detection with integrated channel estimation for coded DS-CDMA," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1217–1223, Aug. 2002.
- [3] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1085–1104, Jul. 2004.
- [4] H. Li, S. M. Betz, and H. V. Poor, "Performance analysis of iterative channel estimation and multiuser detection in multipath DS-CDMA channels," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1981–1993, May 2007.
- [5] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1266–1270, Aug. 2003.
- [6] T. Zemen, C. F. Mecklenbrauker, J. Wehinger, and R. R. Muller, "Iterative joint time-variant channel estimation and multi-user detection for MC-CDMA," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1469–1478, Jun. 2006.
- [7] H. Niu and J. A. Ritcey, "Iterative channel estimation and decoding of pilot symbol assisted LDPC coded QAM over flat fading channels," presented at the Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA, Nov. 2002.
- [8] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-Survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 354–364, Feb. 1995.
- [9] C. N. Georghiadis and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 300–308, Mar. 1997.
- [10] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSDE using the EM algorithm," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1181–1193, Aug. 1999.
- [11] A. Anastasopoulos and K. M. Chugg, "Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1638–1649, Oct. 2000.
- [12] G. K. Kaleb and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2406–2413, Jul. 1994.
- [13] E. Chiavaccini and G. M. Vitetta, "MAP symbol estimation on frequency-flat Rayleigh fading channels via a bayesian EM algorithm," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1869–1872, Nov. 2001.
- [14] M. Nissila and S. Pasupathy, "Adaptive Bayesian and EM-based detectors for frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1325–1336, Aug. 2003.
- [15] S.-H. Wu, U. Mitra, and C.-C. J. Kuo, "Joint channel estimation and multiuser detection for multipath fading channels in DS-CDMA systems," presented at the IEEE Int. Conf. Commun., Anchorage, AK, May 2003.
- [16] S.-H. Wu, U. Mitra, and C.-C. J. Kuo, "Noncoherent multiuser detection of DS-CDMA over multipath fading channels," presented at the IEEE Int. Conf. Commun., Paris, France, Jun. 2004.
- [17] S.-H. Wu, U. Mitra, and C.-C. J. Kuo, "Iterative MAP channel estimation and multiuser detection for DS-CDMA in frequency-selective fading channels," presented at the Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA, Nov. 2004.
- [18] S.-H. Wu, U. Mitra, and C.-C. J. Kuo, "Graph representation for joint channel estimation and symbol detection," presented at the IEEE Globecom, Dallas, TX, Dec. 2004.
- [19] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [20] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.
- [21] A. Kocian and B. H. Fleury, "EM-based joint data detection and channel estimation of DS-CDMA signals," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1709–1720, Oct. 2003.
- [22] J. Wehinger, T. Zemen, A. Kocian, and B. H. Fleury, "Low-complexity joint data detection and channel estimation in time-varying flat-fading channels within the SAGE framework," presented at the NEWCOM-ACORN Joint Workshop, Vienna, Austria, Sep. 2006.
- [23] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
- [24] Q. Li, C. N. Georghiadis, and X. Wang, "Blind multiuser detection in uplink CDMA with multipath fading: A sequential EM approach," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 71–81, Jan. 2004.
- [25] M. J. Borran and M. Nasiri-Kenari, "An efficient technique for CDMA communications systems based on the expectation maximization algorithm," presented at the IEEE Int. Symp. Spread Spectrum Techniques Applicat., Mainz, Germany, Sep. 1996.
- [26] C. Carlemalm, A. Logothetis, and H. V. Poor, "Channel estimation and demodulation of asynchronous CDMA signals in frequency-selection fading channels," presented at the EUSIPCO, Tampere, Finland, Sep. 2000.
- [27] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
- [28] H. Zamiri-Jafarian and S. Pasupathy, "EM-based recursive estimation of channel parameters," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1297–1302, Sep. 1999.
- [29] S. Lam, K. Plataniotis, and S. Pasupathy, "Isometric data sequences and data modulation schemes in fading," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 406–415, Mar. 2004.
- [30] G. Colavolpe, G. Ferrari, and R. Raheli, "Noncoherent iterative (turbo) decoding," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1488–1498, Sep. 2000.
- [31] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.



- [32] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 941–950, Feb. 1994.
- [33] A. Logothetis and V. Krishnamurthy, "Expectation maximization algorithms for MAP estimation of jump markov linear systems," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2139–2156, Aug. 1999.
- [34] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 4, pp. 477–489, Apr. 1988.
- [35] A. Anastasopoulos and K. M. Chugg, "An efficient method for simulation of frequency selective isotropic Rayleigh fading," presented at the IEEE Vehicular Technol. Conf., Phoenix, AZ, May 1997.
- [36] S.-H. Wu, U. Mitra, and C.-C. J. Kuo, "Performance of linear reduced-rank multistage receivers for DS-CDMA in frequency-selective fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3493–3517, Oct. 2005.
- [37] L. Lindbom, A. Ahlén, M. Sternad, and M. Falkenström, "Tracking of time-varying mobile radio channels. II. A case study," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 156–167, Jan. 2002.



**Sau-Hsuan Wu** (S'03–M'04) received the B.S. and M.S. degrees, both in engineering science, from the National Cheng Kung University, Taiwan, R.O.C., in 1990 and 1993, respectively. From 1993 to 1995, he served in the Army of Taiwan and from 1995 to 1999, he worked as a circuit and system design engineer in Taiwan. In 2003, he received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles.

From 2003 to 2004, he was a postdoctoral research fellow at the Department of Electrical Engineering, University of Southern California. From 2004 to 2005, he served as a technical consultant for Winbond Electronics Corporation America, developing wireless MIMO-OFDM products. Since 2005, he has been an Assistant Professor in the Department of Communication Engineering, National Chiao Tung University, Taiwan. His research interest lies in the application of signal processing theory to wireless communication systems.

**Urbashi Mitra** (S'88–M'88–SM'04–F'07) received the B.S. and M.S. degrees, both in electrical engineering and computer science, from the University of California at Berkeley in 1987 and 1989, respectively. From 1989 to 1990, she worked as a Member of Technical Staff at Bellcore, Red Bank, NJ. In 1994, she received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ.

From 1994 to 2000, she was a member of the faculty of the Department of Electrical Engineering, Ohio State University, Columbus. In 2001, she joined the Department of Electrical Engineering, University of Southern California, Los Angeles, where she is currently a Professor. She served as co-Director of the Communication Sciences Institute at the University of Southern California from 2004 to 2007.

Dr. Mitra is currently an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY and the *Journal of Oceanic Engineering*. She was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1996

to 2001. She served as a member of the IEEE Information Theory Society's Board of Governors from 2002 to 2007. She is the recipient of the 2001 Okawa Foundation Award, 2000 Lumley Award for Research (OSU College of Engineering), 1997 MacQuigg Award for Teaching (OSU College of Engineering), 1996 National Science Foundation (NSF) CAREER Award, 1994 NSF International Post-doctoral Fellowship, 1998 Lockheed Leadership Fellowship, and 1987 California Microelectronics Fellowship. She co-chaired the IEEE Communication Theory Symposium at ICC 2003 in Anchorage, AK, and the first ACM Workshop on Underwater Networks at Mobicom 2006, Los Angeles, CA. She was the tutorials Chair for IEEE ISIT 2007 in Nice, France, and is currently the Finance Chair for IEEE ICASSP 2008 in Las Vegas, NV. She has held visiting appointments at: the Eurecom Institute, Rice University, and Stanford University. She was a Texas Instruments Visiting Professor in Fall 2002 at Rice University.



**C.-C. Jay Kuo** (S'83–M'86–SM'92–F'99) received the B.S. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He was Computational and Applied Mathematics (CAM) Research Assistant Professor in the Department of Mathematics at the University of California, Los Angeles, from October 1987 to December 1988.

Since January 1989, he has been with the Department of Electrical Engineering-Systems and the Signal and Image Processing Institute at the University of Southern California, Los Angeles, where he currently has a joint appointment as Professor of electrical engineering and mathematics. His research interests are in the areas of digital signal and image processing, audio and video coding, multimedia communication technologies and delivery protocols, and embedded system design. He has guided about 85 students to their Ph.D. degrees and supervised 20 postdoctoral research fellows. He has delivered more than 360 invited lectures in conferences, research institutes, universities and companies. He is co-author of about 140 journal papers, 740 conference papers and 9 books.

Dr. Kuo is a Fellow of IEEE and SPIE and a member of SIAM and ACM. He is Editor-in-Chief for the *Journal of Visual Communication and Image Representation*, Associate Editor for IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING and Editor for the *Journal of Information Science and Engineering* and the *EURASIP Journal of Applied Signal Processing*. He is also on the Editorial Board of the IEEE SIGNAL PROCESSING MAGAZINE. He served as Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING in 1995–1998 and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 1995–1997. Dr. Kuo received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively. He received the best paper award from the multimedia communication Technical Committee of the IEEE Communication Society in 2005, the best student paper award from the IEEE Vehicular Technology Fall Conference (VTC-Fall) in 2006 and the best paper award from the IEEE Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP) in 2006. He was an IEEE Signal Processing Society Distinguished Lecturer in 2006.