Joint Data Detection and Estimation of Time-Varying
Multipath Rayleigh Fading Channels in Asynchronous
DS-CDMA Systems with Long Spreading Sequences

PEI XIAO
School of Electrical, Electronic
and Computer Engineering
University of Newcastle
NE1 7RU, United Kingdom
pei.xiao@ncl.ac.uk

ERIK G. STRÖM
Communication Systems Group
Dept. of Signals and Systems
Chalmers Univ. of Technology
SE-412 96 Göteborg, Sweden
erik.strom@chalmers.se

Abstract. In this paper, we present a joint approach to data detection and channel estimation for the asynchronous
direct-sequence code-division multiple access (DS-CDMA) systems employing orthogonal signaling formats and long
scrambling codes. Our emphasis is placed on different channel estimation algorithms since the performance of a com-
munication system depends largely on its ability to retrieve an accurate measurement of the underlying channel. We
investigate channel estimation algorithms under different conditions. The estimated channel information is used to en-
able coherent data detection to combat the detrimental effect of the multiuser interference and the multipath propagation
of the transmitted signal. In the considered multiuser detector, we mainly use interference cancellation techniques, which
are suitable for long-code CDMA systems. Interference cancellation and channel estimation using soft estimates of the
transmitted signal is also proposed in this paper. Different channel estimation schemes are evaluated and compared in
terms of mean square error (MSE) of channel estimation and bit error rate (BER) performance. Based on our analysis
and numerical results, some recommendations are made on how to choose appropriate channel estimators in practical
systems.

1 INTRODUCTION

A CDMA cellular communication system is inherently interference-limited. This is due to the difficulty of main-
taining orthogonality on the reverse link between channels used by independent mobile stations, which transmit asyn-
chronously. This form of interference limits the uplink capacity severely. In addition to multiple access interfer-
ence (MAI), CDMA systems also suffer from multipath fading. Mobile radio communication channels are time-
varying channels characterized by the presence of both delay and Doppler spread. Depending on the delay spread
and the data rate, the channel may be approximately flat fading or frequency-selective fading. The latter one pro-
duces intersymbol interference (ISI), since the received signal includes multiple delayed and attenuated versions of the
transmitted waveform.

Accurate knowledge (or a good estimate) of the underlying channel is essential for mitigating interference, ISI,
and fading. If the channel estimates are not reliable, the performance of algorithms, such as multiuser detectors and
coh erent Rake receivers, degrade significantly. Channel estimation consequently is an important issue in mobile com-
munications and good channel estimates have a very important impact on the overall performance of the system.

Several channel estimators, e.g., subspace-based estimators, and maximum likelihood estimators have been pro-
posed for short spreading codes in, e.g., [1, 2]. However, current and next generation CDMA systems use long
spreading codes, i.e., codes with periods much larger than the symbol duration. For long-code CDMA, several at-
tempts have been made in obtaining channel estimates. In [3], a subspace-based algorithm for blind channel es-
timation of a synchronous CDMA downlink was proposed. It was shown that the estimation accuracy can be increased
considerably using a decision feedback approach. However, a time invariant multipath channel was assumed in [3].
The time-varying nature of the fading channel prohibits the use of subspace algorithms, since the received signal is not constrained to any particular subspace if channel parameters are time-varying.

The estimation of channel parameters in a DS-CDMA system with \( M \)-ary orthogonal modulation, which is the main concern of this paper, has been the subject of study in several papers (see, for instance, [4, 5, 6, 7]). A maximum likelihood (ML) channel estimator was employed in [4] for estimating time-varying multipath channels. In [5], a blind channel estimation strategy based on an adaptive Wiener filtering approach that yields unbiased channel estimates and low estimation variance was proposed. Joint channel estimation and data detection has been treated in [6, 8, 9, 7, 10, 11, 12]. A multistage decision-directed channel estimation scheme for a coherent DS-CDMA system with \( M \)-ary orthogonal signaling was proposed in [6]. The channel estimates are obtained by using the decisions from the detector. The correlator outputs are averaged to improve the accuracy of the channel estimates, and the detector uses these estimates for coherent detection of the transmitted symbols. Two efficient iterative structures for joint multiuser detection and channel estimation were presented in [8]. The schemes resulted from an application of the expectation-maximization and the space-alternating generalized expectation-maximization algorithms. Some suboptimum detectors, e.g., the symbol-by-symbol detector and the per-survivor sequence detector were proposed in [9] to overcome the prohibitive computational complexity imposed by the optimum multiuser detector. Kalman filters are used to track the fading channel for diversity combining and the trellis updates. Adaptive versions of these detectors employing recursive-least-square minimum-mean-square-error filter were also introduced. They only require the knowledge of the signature of the user of interest and can be implemented using systolic arrays to exploit parallel signal processing computation. In [7, 10], the estimated channel parameters are used by an interference canceler to achieve successively improved receiver performance. This idea was extended to a convolutionally coded system in [11, 12], where soft-decision feedback is used for interference cancellation to reduce the effect of error propagations.

In this paper, we adopt an iterative scheme for joint channel estimation and multiuser detection, similar to the idea presented in the above references. However, in contrast to the previous work, we take a systematic approach and investigate channel estimation algorithms under different conditions using both classical and Bayesian methods. First, we show that the maximum likelihood (ML) algorithm is efficient when the underlying channels are completely unknown and when the fading processes are slow. With the knowledge of the noise and channel statistics, we show that employing a linear MMSE estimator can yield more noise-resistant estimates. Other alternatives, like first and second order Kalman filters, which take into account the correlative nature of the Rayleigh fading channels, are also introduced. All the channel estimators are decision-directed and can work in conjunction with coherent data detection (interference cancellation in particular). To prevent error propagation from the decision feedback, channel estimation and interference cancellation using soft information is also proposed in this paper. The proposed soft scheme differs from the one proposed in [11, 12] in that the latter is designed for the systems employing a convolutional code. In this case, the soft information is readily available from the soft-output channel decoder. For the uncoded systems, like the one considered in this work, the soft information has to be derived by some other means. To this end, we proposed a maximum a posteriori (MAP) based soft interference cancellation scheme, which is another contribution of this paper.

The remainder of the paper is organized as follows. In Section 2, the system model is presented. Channel estimation and interference cancellation algorithms are introduced in Section 3. Soft interference cancellation and channel estimation are proposed in Section 4. Different algorithms are thoroughly assessed and compared through extensive computer simulations in Section 5. Finally, some conclusions are drawn in Section 6. The derivation of the Cramer-Rao lower bound (CRLB) for the ML estimator and the error covariance for the LMMSE estimator are given in the Appendix.

2 System Model

Some notational conventions used in this paper are as follows. The transpose, conjugate transpose, and 2-norm of a vector \( x \) are denoted by \( x^T \), \( x^* \), and \( \|x\| = \sqrt{x^*x} \), respectively. The \( n^{th} \) element of a vector \( x \) is denoted by \( x_n \). The symbols \( \mathbb{R} \) and \( \mathbb{C} \) denote the real field and complex field, respectively.

The system under study is an asynchronous DS-CDMA system. The employed modulation scheme is a special case of the orthogonal multi-pulse modulation introduced in [13] and is implemented with the Walsh (Hadamard) code. The system resembles the uplink of an IS-95 system in that the narrow-band bit stream is spread by one of \( M \) possible Walsh codewords, which are not used for user separation, but for \( M \)-level modulation. The transmitted chip sequence from a particular user is the concatenation of one of \( M \) possible Walsh sequences (representing the transmitted symbol) and a long scrambling code. Fig. 1 is the block diagram of the transmitter showing the signal path for the \( k^{th} \) user, which is very similar to the system model used in [4]. The \( k^{th} \) user’s \( j^{th} \) symbol is denoted by \( i_k(j) \in \{0, 1, \ldots, M-1\} \), and mapped into \( w_k(j) \in \{w_0, \ldots, w_{M-1}\} \), which is one of the \( M \) orthogonal signal alternatives. The Walsh codeword
where $n(t)$ has the second moments $E[n(t)n(s)] = 0$ and $E[n(t)n^*(s)] = N_0\delta(t-s)$, and $\delta(\cdot)$ is the Dirac delta function. The average power of $h_{k,l}(t)$ is denoted by $P_{k,l} = E[|h_{k,l}(t)|^2]$.

The output from the chip-matched filter, denoted by $y_t = r(t) \ast \psi(-t)$, is sampled every $T_c$ seconds to yield

$$y(iT_c) = r(iT_c) \ast \psi(-t)|_{t=iT_c} = \nu(iT_c) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} h_{k,l}(t-\tau_{k,l})a_k(n) \cdot \psi(t-nT_c-\tau_{k,l}) \ast \psi(-t)|_{t=iT_c}$$

where $\nu(t) = n(t) \ast \psi(-t)$ and the noise sample $\nu(iT_c)$ is a zero-mean complex Gaussian random variable with second moments $E[|\nu(iT_c)|^2] = N_0$ and $E[|\nu^2(iT_c)|] = 0$.

Let $p_{k,l}$ and $\epsilon_{k,l} \in [0, 1]$ be the integer and fractional part of the delay $\tau_{k,l}$, i.e., $\tau_{k,l} = (p_{k,l} + \epsilon_{k,l}) T_c$. Assuming that the channel gains are approximately constant during one symbol duration, the vector $r(k,j) \in \mathbb{C}^{N_k}$ corresponding to the $k$th user’s $j$th symbol contains $N_k = N + p_{k,L_k} - p_{k,1}$ samples of $y(iT_c)$ and can be written in the following forms

$$r(k,j) = A(k,j)h(j) + n(k,j)$$

$$= X_{k,i_1}^{i_1}(j)h_1(j) + \text{MAI}(k,j) + \text{MAI}(k,j) + n(k,j)$$

As shown in Fig. 2, $y(iT_c)$ consists of contributions from all users’ path signals and the additive noise. The $n(k,j)$ vector is a vector of the noise samples $\nu(iT_c)$. Each column of the matrix $A(k,j)$ represents the contribution from each path and is essentially the product of the channel gain and a shifted version of the appropriate user’s chip sequence (the shift is due to the path delay). The columns of $A(k,j)$ are weighted together by $h_1(j)$, whose elements are the path gains of all users’ paths. From Fig. 2, we see that $r(k,j)$ can also be written as the sum of four terms: the signal of interest, the intersymbol interference (ISI), the multiple access interference (MAI), and the noise. The signal of interest is the part of $y(iT_c)$ that is due to the $k$th user’s $j$th symbol. In Fig. 2, the signal of interest for the first user, $X_{k,i_1}^{i_1}(j)h_1(j)$ is marked with bold lines. The columns of the matrix $X_{k,i_1}^{i_1}(j)$ are essentially shifted versions of the chips due to the $k$th user’s $j$th symbol, one column per
Figure 2: Sketch of contributions from the users' paths to the received signal. The signal during the indicated time interval is represented by \( r(1,j) \).

path. The columns of \( X_{k,1}(j) \) are weighted together by the vector \( h_k(j) \), whose elements are the path gains of the \( k \)th user’s paths. The contribution only from the \( k \)th user’s \( j \)th symbol can be written as \( X_{k,1}(j)h_k(j) \).

The matrix \( A(k,j) \in \mathbb{R}^{N_k \times L_{\text{tot}}} \) (\( L_{\text{tot}} \) is the total number of paths of all users, i.e., \( L_{\text{tot}} = \sum_{k=1}^{K} L_k \)) is defined as

\[
A(k,j) = \begin{bmatrix}
A_1(k,j) & \cdots & A_K(k,j)
\end{bmatrix},
\]

\[
A_i(k,j) = [a_{i,1}(k,j) \cdots a_{i,L_i}(k,j)],
\]

\[
[a_{i,l}(k,j)]_n = (1 - \epsilon_i, l)a_i(jN + p_{k,1} + n - p_{i,l}) + \epsilon_i, l a_i(jN + p_{k,1} + n - p_{i,l} - 1)
\]

where \( L_{\text{tot}} = \sum_{k=1}^{K} L_k \). Note that \( A_1(k,j) = X_{1,1}(j) + \text{ISI}(k,j) \), i.e., \( X_{1,1}(j) \) is the part of \( A_1(k,j) \) that only contains the desired signal, see Fig. 2. The channel vector \( h(j) \in \mathbb{C}^{L_{\text{tot}}} \) is defined as

\[
h(j) = [h^T_1(j) \ h^T_2(j) \ \cdots \ h^T_K(j)]^T,
\]

\[
h_i(j) = [h_{i,1}(jT) \ h_{i,2}(jT) \ \cdots \ h_{i,L_i}(jT)]^T.
\]

The matrix \( X_{k,m}(j) \) is defined as

\[
X_{k,m}(j) = [x_{k,1,m}(j) \ x_{k,2,m}(j) \ \cdots \ x_{k,L_k,m}(j)]
\]

\[
x_{k,l,m}(j) = (1 - \epsilon_{k,l}) \begin{bmatrix}
0_{(p_{k,1}, p_{k,1} + 1)} & C_k(j)w_m
\end{bmatrix}
\]

\[
+ \epsilon_{k,l} \begin{bmatrix}
0_{(p_{k,1}, p_{k,1} + 1)} & C_k(j)w_m
\end{bmatrix}
\]

where \( w_m \) is the \( m \)th column of the \( N \times N \) Hadamard matrix. The vector \( x_{k,l,m} \) contains the transmitted chip sequence due to the \( k \)th user’s \( j \)th symbol from the \( l \)th path based on the hypothesis that the \( n \)th Walsh symbol is transmitted.

For simplicity of notation, we will suppress the index \( k \) and/or \( j \) from \( r(k,j), A(k,j), x(j), X_{k,1}(j) \) and \( h_k(j) \), etc., whenever no ambiguity arise.

### 3 Estimation of Fading Processes

The task of a channel estimator is to estimate the fading vector \( h \) given the received observation \( r \) and some knowledge of the transmitted data. Depending on the form of the data knowledge, channel estimation can be either decision-directed or pilot-aided. The former utilizes decisions on the transmitted data to form estimates of the transmitted signals \( A \), which in turn are used to estimate the channel coefficients. The second approach makes the use of pilot symbols (\( A \) is known in this case). The use of pilots simplifies channel estimation with the penalty of wasting channel resources.

In this paper, we focus on the first approach and make an extensive investigation on different alternatives for estimating time-varying multipath Rayleigh fading channels in absence of pilot symbols. All considered channel estimators are iterative and decision-directed and can work in conjunction with coherent data detection. The estimation procedure at the \( n \)th iteration uses the data estimates from the previous stage.

Depending on whether some prior knowledge about the parameters to be estimated is used or not, estimation can be conducted with a Bayesian or a classical approach. Both approaches are discussed in this paper. A maximum likelihood-based channel estimator is derived based on the classical approach, in which the channel vector \( h \) is assumed to be deterministic and unknown. In a Bayesian approach, \( h \) is assumed to be a random vector whose particular realization needs to be estimated. The estimation accuracy can then possibly be improved by exploiting some prior knowledge about \( h \). The Bayesian philosophy leads,
in this paper, to the linear MMSE estimator and the Kalman filters for channel estimation.

### 3.1 Maximum likelihood channel estimator

Given an estimate of the data matrix \( \hat{A} \) in (1), then \( h \) can be estimated as \( \hat{h} = \hat{A}^\dagger r \), where \( \hat{A}^\dagger = (\hat{A}^* \hat{A})^{-1} \hat{A}^* \) denotes the left pseudo-inverse of \( \hat{A} \) (assuming \( \hat{A} \) has full column rank) [4]. The estimator is the maximum likelihood estimator of \( h \) if \( \hat{A} \) is known (i.e., if \( \hat{A} = A \)) or if \( \hat{A} \) is the ML estimate of \( A \). Although technically incorrect, we will call \( \hat{A}^\dagger r \) the ML estimate of \( h \), regardless of how \( \hat{A} \) is obtained.

In case of correct decisions, i.e., \( \hat{A} = A \), then \( \hat{h} = A^\dagger r = \hat{A}^\dagger (Ah + n) = h + A^\dagger n \), which is an unbiased estimate of \( h \). In Appendix A.1, it is shown that the ML estimator with \( \hat{A} = A \) is an efficient estimator which attains the Cramer-Rao lower bound (CRLB).

This procedure will suffer from a so-called dimensionality problem. When the total number of paths of all the users is greater than the number of chips in the vector \( r(k, j) \), i.e., \( L_{\text{tot}} \geq N \) (\( L_{\text{tot}} = \sum_{k=1}^{K} L_k \)), the matrix \( \hat{A} \) will not have full column rank and the above-mentioned procedure will become useless. The problem can be resolved by stacking several \( r(j) \) vectors on top of each other and assuming that the channel remains static during several symbol intervals. In particular, suppose \( h(j) \approx h(j + 1) \), we can then write

\[
\begin{bmatrix}
  r(j) \\
  r(j + 1)
\end{bmatrix} =
\begin{bmatrix}
  \hat{A}(j) & 0 \\
  0 & \hat{A}(j + 1)
\end{bmatrix}
\begin{bmatrix}
  h(j) \\
  h(j + 1)
\end{bmatrix} +
\begin{bmatrix}
  n(j) \\
  n(j + 1)
\end{bmatrix}
\approx
\begin{bmatrix}
  \hat{A}(j) \\
  \hat{A}(j + 1)
\end{bmatrix} h(j) +
\begin{bmatrix}
  n(j) \\
  n(j + 1)
\end{bmatrix}
\]

The ML channel estimation algorithm can then be reformulated, following [4], as

\[
\hat{h}^{\text{ML}}(j) = \begin{bmatrix}
  \hat{A}(j) \\
  \hat{A}(j + 1)
\end{bmatrix}^\dagger
\begin{bmatrix}
  r(j) \\
  r(j + 1)
\end{bmatrix}
\]

which will produce usable estimates as long as \( 2N_k > L_{\text{tot}} \). Obviously, this scheme can be extended further by stacking several \( r(k, j) \) vectors on top of each other. In the derivation of CRLB in Appendix A.1, we see that stacking also has the effect of noise averaging and tends to reduce the error of the channel estimation. However, for relatively fast fading channels, the stacking may have opposite effect and reduce the quality of the estimates.

### 3.2 Linear MMSE Channel Estimator

The linear MMSE estimate of \( h \) is \( W^* r \), where \( W \) is chosen to minimize the mean square error \( E[\|h - W^* r\|^2] \). The optimum matrix of \( W \) under the MMSE criterion can be computed as

\[
W_{\text{LMMSE}} = \arg \min_W E[\|h - W^* r\|^2] = R^{-1} \Phi
\]

\[
R = E[rr^*] = E[(Ah + n)(Ah + n)^*] = APA^* + N_0 I
\]

\[
\Phi = E[hh^*] = E[(Ah + n)h^*] = AP
\]

\[
P = E[hh^*] = \text{diag}(P_{k,1}, P_{k,2}, \ldots, P_{k,L_k})
\]

where \( P_{k,l} \) is the average received power from the \( k^{th} \) user’s \( l^{th} \) path. Combining the above equations, the linear MMSE estimate of \( h \) can be formulated as

\[
\hat{h}^{\text{LMMSE}} = W_{\text{LMMSE}}^* r = \Phi^* R^{-1} r
= P^* A^* (A P A^* + N_0 I)^{-1} r
\]

It is shown in Appendix A.2 that the Bayesian Gauss-Markov Theorem leads to the same MMSE estimator. Furthermore, the estimation covariance matrix is also derived there.

From (4), we see that the signal and noise power level \( P_{k,l} \) and \( N_0 \) must be known or estimated to carry out LMMSE channel estimation. It is worth noticing that \( A \) does not need to be full rank to ensure that \( (APA^* + N_0 I) \) is invertible. Since we use a decision-directed approach for channel estimation, the transmitted data \( A \) is unknown and has to be replaced by its estimate \( \hat{A} \) in (4).

The fading processes are low-pass in nature with bandwidths that are determined by the Doppler frequency. We should therefore be able to improve the estimates by low-pass filtering (smoothing). A simple smoothing procedure is to feed \( \hat{h}(j) \) through an FIR filter with impulse response \( g(n) \) [4], which yields the smoothed channel gain vector \( \tilde{h}(j) \) as

\[
\tilde{h}(j) = \sum_{k=j-N_s}^{j+N_s} \hat{h}(k) g(j - k)
\]

Experiments indicate that exact form of the filter is not crucial. In our simulations, we use a smoothing filter derived from a Hamming window of length \( 2N_s + 1 = 19 \) or 9 (for slow or fast fading channels), normalized such that \( \sum_{k=1}^{N_s} g(k) = 1 \). As will be evidenced by numerical results, smoothing really improves the ML and LMMSE channel estimates, which subsequently leads to more reliable data detection.

### 3.3 First-order Kalman filter

If \( h \) and its statistics are completely unknown, the ML estimator is the best we can do. However, we know that channel gains are correlated in time. One way to exploit this knowledge is to use a smoothing filter to improve estimation results as introduced above. An alternative approach is to consider \( h \) to be a realization of a random process when developing of channel estimation algorithms. This leads to the use of, e.g., Kalman filters, which are widely used for estimating random parameters that evolve in time according to some dynamic
model [14]. Kalman filters have been used for channel estimation, e.g., in [9, 15, 16, 17, 18]. Here, we extend the use of Kalman filters to estimation of multipath fading channels for CDMA systems with orthogonal modulation. The idea is similar to the one presented in [9], i.e., the joint data detection and channel estimation using Kalman filters. However, our iterative scheme is block based, i.e., at each iteration, both data detection and channel estimation are carried out for the whole block of data, and the performance is improved as iteration goes on; whereas the Kalman filtering in [9] is conducted on symbol-by-symbol basis, and there is no multistage iteration involved in channel estimation and data detection. Furthermore, different detection techniques are used in [9], e.g., decorrelator with multipath combining or an approximate maximum a posteriori (MAP) sequence detection as well as their adaptive implementations.

We assume that the channels are independent Rayleigh fading channels with Clarke’s power spectral density [19], illustrated in Fig. 3 and given by

$$S(f) = \begin{cases} \frac{K}{\sqrt{(1-(f/f_d)^2)^2}}, & |f| < f_d \\ 0, & |f| \geq f_d \end{cases}$$  \hspace{1cm} (6)$$

where $f_d$ is the maximum Doppler frequency. The channel gain $h_{k,l}(t)$ is a complex circular Gaussian process with autocorrelation function $E[h_{k,l}(t)h_{k,l}(t+\tau)] = P_{k,l}J_0(2\pi f_d\tau)$, where $J_0(x)$ is the zeroth order Bessel function of the first kind. Based on the relation $\int_{f_d}^{-f_d} S(f) df = P_{k,l}$, we can readily derive the constant $K = P_{k,l}/(\pi f_d)$.

A random process with power spectral density as expressed in (6) can be approximated by a finite order autoregressive (AR) process. The correlation between channel coefficients can simply be modeled by the first-order vector Gauss-Markov model as $h(t) = Fh(t-1) + u(t)$, where $F \in \mathbb{R}^{L_{tot}\times L_{tot}}$ is the state transition matrix and $u(t) \in \mathbb{C}^{L_{tot}}$ is the driving noise vector (WGN sequence) with $E[u(t)] = 0$ and covariance matrix $Q$. The current input $h(t)$ depends only on the state of the system at the previous symbol time $h(t-1)$ and the current input $u(t)$. The state $h(t-1)$ accumulates the effect of all past inputs to the system.

Since different paths are uncorrelated with each other, we can let $F$ and $Q$ be diagonal matrices, i.e., $F = \text{diag}(f_{1,1}, f_{1,2}, \ldots, f_{K,1}, \ldots, f_{K,L_{tot}})$ and $Q = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{K,1}^2, \ldots, \sigma_{K,L_{tot}}^2)$. Using the notation $h_{k,l}[j] = h_{k,l}(jT)$, the vector Gauss-Markov model becomes $L_{tot}$ independent scalar models: $h_{k,l}[j] = f_{k,l}h_{k,l}[j-1] + u_{k,l}[j]$. By solving the equations

$$E[h_{k,l}[j]h_{k,l}[j-1]] = f_{k,l}P_{k,l} + E[u_{k,l}[j]h_{k,l}[j-1]]$$

$$E[h_{k,l}[j]^2] = f_{k,l}^2E[h_{k,l}[j-1]^2] + E[u_{k,l}[j]^2]$$

$$= f_{k,l}^2P_{k,l} + \sigma_{k,l}^2$$

we obtain the coefficients for the first-order AR model

$$f_{k,l} = \frac{E[h_{k,l}[j]h_{k,l}[j-1]]}{P_{k,l}} = J_0(2\pi f_dT)$$

$$\sigma_{k,l}^2 = P_{k,l}(1 - f_{k,l}^2)$$

Note that the maximum Doppler frequency $f_d$ can differ from path to path. However, in the special case when all the paths from all users have the same received power and Doppler frequency, the matrices $F$, $Q$ are simplified to $F = fI$, $Q = \sigma_0^2I$. A training method was proposed in [15] to estimate the state transition matrix $F$, and an algorithm for tracking Doppler shifts can be found in [17]. Here, we focus on the estimation of complex channel gains and assume that the Doppler shift is known after a training phase.

Recall that our observation vector (measurement model) is, from (1), $r(j) = A(j)h(j) + n(j) \in \mathbb{C}^N$, where $n(j) \sim \mathcal{CN}(0, C)$ and $C = N_0I_N$. The channel estimate based on the previous received observations, $\hat{h}(j|j-1)$, can be obtained with the following Kalman recursive estimation procedure [14]

$$\hat{h}(j|j-1) = F\hat{h}(j-1|j-1)$$

$$M(j|j-1) = FM(j-1|j-1)F^T + Q$$

$$K(j) = M(j|j-1)A(j)^T[C + A(j)M(j|j-1)A(j)^T]^{-1}$$

$$\hat{h}(j|j) = \hat{h}(j|j-1) + K(j)[r(j) - A(j)\hat{h}(j|j-1)]$$

$$M(j|j) = (I - K(j)A(j))M(j|j-1)$$  \hspace{1cm} (7)$$

From the above equation, we see that the minimum Bayesian MSE, $M(j|j-1)$, is computed as an integral part of the estimator. The performance measure of the Kalman filter is therefore different from the estimators discussed earlier. In our experiments, the recursion is initialized by $\hat{h}(-1|1) = 0$ and $M(-1|1) = 100I$, which were chosen to reflect little knowledge of the initial state. Experiments indicate that the exact initial value for $M(-1|1)$...
is not crucial. The Bayesian MSE will decrease gradually as the algorithm converges. From (7), we see that the inversion of an \( N \times N \) matrix is required to find the Kalman gain \( K(j) \).

### 3.4 Second-order Kalman filter

The Kalman filter employed above is based on modeling the fading process as a first-order AR process. Further improvement in the Kalman channel estimator will be possible, with increased complexity, if a higher order AR model is used. For example, the correlation between channel coefficients can be more closely approximated by

\[
h(j) = F_1 h(j-1) + F_2 h(j-2) + u(j) \quad (8)
\]

As shown in Fig. 4, the second-order AR model expressed by (8) yields a better approximation of the Doppler spectrum (shown in Fig. 3) than the first-order AR model. Consequently, the accuracy of the Kalman filter should improve. The second-order vector Gauss-Markov model and measurement model become

\[
\begin{bmatrix}
  h(j-1) \\
  h(j)
\end{bmatrix} =
\begin{bmatrix}
  0 & 1 \\
  F_2 & F_1
\end{bmatrix}
\begin{bmatrix}
  h(j-2) \\
  h(j-1)
\end{bmatrix}
+ \begin{bmatrix}
  u(j)
\end{bmatrix}
\]

\[
\begin{bmatrix}
  r(j-1) \\
  r(j)
\end{bmatrix} =
\begin{bmatrix}
  A(j-1) & 0 \\
  0 & A(j)
\end{bmatrix}
\begin{bmatrix}
  h(j-1) \\
  h(j)
\end{bmatrix}
+ \begin{bmatrix}
  n(j-1) \\
  n(j)
\end{bmatrix}
\]

For simplicity, we assume all the paths have equal received power and the same Doppler frequency in the following derivation. However, the extension to unequal power and Doppler frequency is straightforward. Now, the covariance matrix \( Q = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_q^2 \end{bmatrix} \) the state of the system at the previous symbol time is \( \begin{bmatrix} h(j-2) \\ h(j-1) \end{bmatrix} \), and the state transition matrix is \( F = \begin{bmatrix} 0 & 1 \\ F_2 & F_1 \end{bmatrix} \). The covariance matrix of noise vector \( \begin{bmatrix} n(j-1) \\ n(j) \end{bmatrix} \) becomes \( C = N_0 I_{2N} \).

Since different paths are uncorrelated with each other, we can assume \( F_1 = f_1 I, \ F_2 = f_2 I \), and \( \{f_1, f_2, \sigma_q^2\} \) can be determined by the relationship \( h_{k,l}[j] = f_1 h_{k,l}[j-1] + f_2 h_{k,l}[j-2] + u_{k,l}[j] \). Let us denote

\[
\begin{align*}
E \{ h_{k,l}[j] h_{k,l}[j-1] \} &= J_0(2\pi f_d T) P_{k,l} = J_1 P_{k,l} \\
E \{ h_{k,l}[j] h_{k,l}[j-2] \} &= J_0(4\pi f_d T) P_{k,l} = J_2 P_{k,l}
\end{align*}
\]

The model parameters can then be calculated from

\[
\begin{align*}
E \{ h_{k,l}[j] h_{k,l}[j-1] \} &= f_1 E \{ |h_{k,l}[j-1]|^2 \} + f_2 E \{ h_{k,l}[j-2] h_{k,l}[j-1] \} + E \{ u_{k,l}[j] h_{k,l}^* [j-1] \} \\
\implies J_1 &= f_1 + f_2 J_1; \\
E \{ h_{k,l}[j] h_{k,l}[j-2] \} &= f_1 E \{ h_{k,l}[j-1] h_{k,l}^*[j-2] \} + f_2 E \{ |h_{k,l}[j-2]|^2 \} + E \{ u_{k,l}[j] h_{k,l}^*[j-2] \} \\
\implies J_2 &= f_1 J_1 + f_2; \\
E \{ |h_{k,l}[j]|^2 \} &= f_1^2 E \{|h_{k,l}[j-1]|^2 \} + f_2^2 E \{|h_{k,l}[j-2]|^2 \} + E \{ |u_{k,l}[j]|^2 \} \\
&+ f_1 f_2 E \{ h_{k,l}[j-1] h_{k,l}^*[j-2] + h_{k,l}^*[j-1] h_{k,l}[j-2] \}
\end{align*}
\]

By solving the above equations, we obtain

\[
\begin{align*}
f_2 &= \frac{J_1^2 - J_2}{J_1^2 - 1}; \\
f_1 &= \frac{J_2 - f_2}{J_1}, \\
\sigma_q^2 &= (1 - f_1^2 - f_2^2 - 2 f_1 f_2 J_1) P_{k,l}
\end{align*}
\]

The procedure stated in (7) also applies to the second-order Kalman filter with the following replacements:

\[
\begin{align*}
A(j) &\mapsto \begin{bmatrix} A(j-1) & 0 \\ 0 & A(j) \end{bmatrix}; \\
F &\mapsto \begin{bmatrix} 0 & I_N \\ F_2 & F_1 \end{bmatrix}; \\
Q &\mapsto \begin{bmatrix} 0 & 0 \\ 0 & \sigma_q^2 I \end{bmatrix}; \\
h(j) &\mapsto \begin{bmatrix} h(j-1) \\ h(j) \end{bmatrix}
\end{align*}
\]

The complexity increase by implementing the second-order Kalman filter instead of the first-order is mainly due to the inversion of a \( 2N \times 2N \) matrix required to compute the Kalman gain.

In addition to the estimation of time-varying multipath coefficients, it was shown in [16] that the extended Kalman filter (EKF) can be used to estimate code delay. In this case, the observation (measurement) sequence is nonlinear in the state variables (the propagation delays), and the ordinary Kalman filter equations cannot be used. Thus, the observation sequence must be linearized to allow for a recursive estimation procedure using the Kalman filter equations, and the EKF is a practical solution to this problem.
3.5 Data detection with interference cancellation

The purpose of channel estimation is to enable coherent detection and maximum ratio combining (MRC) in the Rake receiver. In this paper, we concentrate on how to utilize the measured channel for detecting the transmitted data and how to use the detected data to enhance the channel estimation. The principle is that the accuracy of channel estimation depends on the accuracy of the data detection, and vice versa. The channel is better estimated when the transmitted data are more accurately detected, the performance is improved by repeating the estimation and detection process in an iterative manner.

The task of the demodulator is to detect the information bits from all users, i.e., detect \( i_k(j) \) for \( k = 1, 2, \ldots, K, \ j = 1, 2, \ldots, L_b \) \( (L_b \) is the block length) given the observation \( r(k, j) \). The decision on the \( k^{th} \) user’s \( j^{th} \) symbol, is found as \( \hat{i}_k(j) = \arg \max_{m \in \{1, 2, \ldots, M\}} z_k(m) \), where \( z_k(m) \) is the decision statistic obtained from the symbol matched filter or multiuser detector (interference canceler in our case).

The conventional detection technique is to form the soft decision by correlating the received signal with the \( M \) possible transmitted waveforms. Without the knowledge of the fading processes, the soft decision is formed in a path-by-path noncoherent manner as

\[
    z_k(m) = \sum_{l=1}^{L_k} |X_k^*_{k,l,m}(j)r(k, j)|^2
\]

Note that (9) is essentially an equal gain combining scheme. It works the best when each path has equal average power, which is assumed in our system. For unbalanced multipath channels, i.e., average power differs from path to path according to the channel statistics, different multipath components should be weighted accordingly based on the mean square value of the channel gain of each path. The conventional receiver has poor performance in multiuser environments since it considers MAI as additive noise, and the knowledge about MAI is not exploited in any way. To compensate for the effect of MAI, multiuser detection techniques can be used. Multistage interference cancellation schemes are known to be simple and effective multiuser detectors for long-code DS-CDMA systems. Iterative schemes using parallel interference cancellation (PIC) and nonlinear MMSE interference cancellation (NICM) for demodulating \( M \)-ary orthogonal signaling formats in DS-CDMA systems were proposed in [4, 5]. For the purpose of this study, we mainly consider the PIC scheme used in [4] and extend this scheme to soft decision based interference cancellation in Section 4. The basic principle is that once the transmitted signals are estimated for all the users at the previous iteration, interference can be removed by subtracting the estimated interference from the received signal \( r(k, j) \) to form the signal vector \( r'(k, j) \), i.e.,

\[
    r' = r - \hat{A}h + X_k\hat{h}_k
\]

where the vector \( \hat{A}h \) represents the estimated contribution from all the users calculated by using the estimated data matrix \( A \) and channel vector \( h \), and the vector \( X_k\hat{h}_k \) is the estimated contribution due to the \( j^{th} \) symbol from all paths of user \( k \). The soft decision with IC can be readily formed as

\[
    z_k(m) = \text{Re}\{h_k^*X_{k,m}^*r'\} \\
    = \text{Re}\{h_k^*X_{k,m}^*[r - \hat{A}h + X_k\hat{h}_k]\} \quad (10)
\]

4 Soft interference cancellation (IC) and channel estimation (CE)

The conventional interference cancellation receiver is subject to performance degradation due to incorrect decisions on interference subtracted from the received signal. To prevent error propagation, soft IC and CE can be used. The rationale is that hard IC and CE tends to increase the interference and propagate errors with incorrect decision feedback; however, with soft cancellation and CE, the soft estimate of an erroneously estimated symbol usually has a small value and does not make much contribution to the feedback error and propagation is therefore hopefully reduced. Soft IC in uncoded \( M \)-ary CDMA system was treated in [5], in which a multistage MMSE linear interference canceler that minimizes the power of residual cancellation error for each user was proposed. In this paper, we propose a MAP-based soft IC scheme and describe how the soft information used for cancellation can be derived. We then further extend the use of the derived soft information for channel estimation.

The soft reliability value for the \( n^{th} \) bit of the Walsh codeword \( w^*_n(j), n = 0, 1, \ldots, M - 1 \) can be directly derived from the received vector \( r(k, j) \). In the following discussion, we use \( M = 8 \) as an example. From Table 1, we know that bits +1 and −1, are equally probable, i.e., \( P(w^*_n(j) = +1) = P(w^*_n(j) = -1) \), for \( n = 1, \ldots, 7 \). A posteriori log-likelihood ratio (LLR) for a transmitted +1 and a transmitted −1 in the bit sequence \{\( w^*_n(j) \)\} given received vector \( r(k, j) \) is defined as [20]

\[
    \lambda(w^*_n(j)) = \ln \frac{f(w^*_n(j) = +1 | r)}{f(w^*_n(j) = -1 | r)} \quad (11)
\]

\[
    = \ln \frac{f(r | w^*_n(j) = +1)P(w^*_n(j) = +1)}{f(r | w^*_n(j) = -1)P(w^*_n(j) = -1)} \quad (11a)
\]

\[
    = \ln \frac{f(r | w^*_n(j) = +1)}{f(r | w^*_n(j) = -1)} \quad (11b)
\]

\[
    = \ln \frac{\sum_{m:w^*_n(j)=+1} f(r | w_m)}{\sum_{m:w^*_n(j)=-1} f(r | w_m)} \quad (11c)
\]

\[
    = \ln m:w^*_n(j)=+1 \sum_{m:w^*_n(j)=-1} f(r | w_m)
\]

\[
    \lambda(w^*_n(j)) = \ln \frac{f(w^*_n(j) = +1 | r)}{f(w^*_n(j) = -1 | r)} \quad (12)
\]

\[
    = \ln \frac{f(r | w^*_n(j) = +1)P(w^*_n(j) = +1)}{f(r | w^*_n(j) = -1)P(w^*_n(j) = -1)} \quad (12a)
\]

\[
    = \ln \frac{f(r | w^*_n(j) = +1)}{f(r | w^*_n(j) = -1)} \quad (12b)
\]

\[
    = \ln \frac{\sum_{m:w^*_n(j)=+1} f(r | w_m)}{\sum_{m:w^*_n(j)=-1} f(r | w_m)} \quad (12c)
\]
where \( f(w^*_k(j) = \pm 1 | r) \) denotes the probability that the \( n \)th bit of the Walsh codeword \( w^*_k(j) \) takes on the value \( \pm 1 \) conditioned on the received vector \( r \), and
\[
f(r | w^*_m) = \pm 1 \) denotes the PDF of the received vector \( r \) conditioned on the bit \( w^*_m(j) = \pm 1 \) is transmitted. We denote \( m : w^*_m(j) = \pm 1 \) as the set of Walsh codes \( \{w_m\} \) that correspond to the code bit \( w^*_m(j) = \pm 1 \). Typically, one term will dominate each sum, which suggests the dual maxima approximation [21]
\[
\lambda(w^*_m) \approx \max_{m : w^*_m(j) = \pm 1} \frac{f(r | w^*_m)}{\max_{m : w^*_m(j) = \pm 1} f(r | w^*_m)}
\]
\[
\lambda(w^*_m) \approx \ln \left( \frac{\max_{m : w^*_m(j) = \pm 1} f(r | w^*_m)}{\max_{m : w^*_m(j) = \pm 1} f(r | w^*_m)} \right)
\]
\[
\lambda(w^*_m) \approx \frac{2}{\max_{m : w^*_m(j) = \pm 1} f(r | s_m)} \quad (12)
\]
where \( X^+ \) denotes the \( k \)-th bit that corresponds to \( \max_{m : w^*_m(j) = \pm 1} f(r | s_m) \), and \( X^- \) is defined similarly. Comparing (12) with (10), one can see that the added complexity by deriving soft values rather than making hard decisions is minor.

Once the LLR value is derived, the soft estimate (expected value given the received observation) for each bit of the Walsh codeword can be computed as
\[
E[w^*_m(j) | r] = (+1) \times P\{w^*_m(j) = +1 | r\} + (-1) \times P\{w^*_m(j) = -1 | r\}
\]
\[
= (+1) \frac{e^{\lambda(w^*_m(j))}}{1 + e^{-\lambda(w^*_m(j))}} + (-1) \frac{e^{-\lambda(w^*_m(j))}}{1 + e^{-\lambda(w^*_m(j))}}
\]
\[
= \tanh(\lambda(w^*_m(j))/2)
\]
\[
E[w^*_m(j) | r] = \tanh(\lambda(w^*_m(j))/2)
\]

The soft estimate \( E[s^q_k] \) for each Walsh chip \( s^q_k, q = 0, \cdots, N - 1 \) is derived by spreading (repetition encoding) the soft bit of Walsh codeword \( E[w^*_m(j) | r] \). The repetition factor is \( N/M \). When \( E(s^q_k | r) \), the soft estimate of the transmitted sequence is available, we can carry out soft IC. The cancellation residual after soft cancellation becomes
\[
r^*_\text{soft} = r - E[y | r] + E[X_k | r] h_k
\]
where \( E[y | r] = E[A | r] h \), and \( E[A | r] \) and \( E[X_k | r] \) are soft estimates of \( A \) and \( X_k \), respectively. The columns of \( E[A | r] \) and \( E[X_k | r] \) are derived by spreading (repetition encoding) the Walsh code \( w^*_m(j) \) with \( C_k \) and compensating for the path delays.

The ML channel estimation expressed by (3) can be reformulated by replacing \( A \) with its soft estimate, which leads to soft version of the ML channel estimator
\[
\hat{h}^\text{ML}_\text{soft}(j) = \left[ E[A(j) | r(j)] \right]^t \left[ E[A(j + 1) | r(j + 1)] \right]
\]
\[
= \tilde{r}(j)^t \left[ E[A(j) | r(j)] \right]^t \left[ E[A(j + 1) | r(j + 1)] \right]
\]
\[
= \hat{r}(j)^t \left[ E[A(j) | r(j)] \right]^t \left[ E[A(j + 1) | r(j + 1)] \right]
\]

Similarly, the soft version of the LMMSE channel estimator expressed by (4), can be formed as
\[
\hat{E}_\text{LMMSE} = \hat{P} \left[ E[A | r](E[A | r])^t \right] \left[ E[A | r] \right] + N_0 I_n^{-1} r
\]

We can also replace \( A[j] \) with \( E[A(j) | r(j)] \) in (7) to obtain the soft versions of the Kalman filters. Comparing (3) versus (14) for the ML estimator, and (4) versus (15) for the LMMSE estimator, one can see that the use of soft information in channel estimation slightly increases the complexity in hardware implementation since \( A \) takes on binary values whereas \( E[A | r] \) takes on continuous values. The derivation of soft values is also little more complicated than making hard decisions, as discussed earlier.

### 5 Numerical Results

In our simulations, each user transmits one of \( M = 8 \) Walsh codes spread to a total length of \( N = 64 \) chips. The effective spreading of the system is \( N/\log_2 M = 64/3 \) chips per information bit. Different users are separated by different scrambling codes \( C_k(j) \) which are random and differ from symbol to symbol. For simplicity, the simulated system is assumed to be chip-synchronous, i.e., all paths delays are assumed to be multiples of \( T_c \). This represents the worst-case interference scenario [22]. However, the system is asynchronous at the symbol level. The normalized Doppler frequency is set to \( f_d T = 0.1 \) in Fig. 9. When the performance of the ML and Kalman filter is examined in the presence of fast fading channels. In the rest of the simulations, \( f_d T = 0.01 \) which represents relatively slow fading channels. Perfect slow power control is assumed in the sense that \( P_k = \sum_{i=1}^{L_k} P_{k,i} \), the average received power, is equal for all users. Different paths are assumed to have equal gains, and the channel coefficients are normalized so that each user has unit received power, i.e.,
$P_{k,1} = P_{k,2} = \cdots = P_{k,L_k}$ and $P_k = \sum_{l=1}^{L_k} P_{k,l} = 1$. The system was simulated for 2-path channels, $L_k = L = 2$ or 3-path channels, $L_k = L = 3$ for all $k$. The spacing between the paths of each user is set to $2T_c$. The number of users is either $K = 8$ or $K = 12$, unless otherwise stated. The simulation results are averaged over random distributions of fading, noise, delay, and scrambling codes through numerous Monte-Carlo runs. The channel estimators are evaluated with respect to the mean square error (MSE) and the resulting BER performance.

Noncoherent matched filters (MF) expressed by (9) were used for the first stage of the PIC scheme to account for the fact that channel estimates are not yet available at the initial iteration. In the following stages, both interference cancellation and channel estimation are carried out in decision-directed mode using the detected data from the previous iteration.

Fig. 5 shows the original fading channel and the results of channel estimation with different schemes. The original ML and LMMSE estimates are noisy. By explicitly taking the correlation of the fading channel into account, Kalman filters result in more reliable channel estimates, especially if the fading channel is modeled as a second-order AR process. We can see from the plots that, after an initial transient, the Kalman filter quickly locks on the true channel values and tracks them closely. The imposition of a correlation constraint prevents the estimate of $h$ from fluctuating too widely in time. The same effect can be achieved by the ML and the LMMSE estimator by an additional low-pass filtering stage, as shown in the plot at the lower right corner of Fig. 5. The quality of the estimated channel is greatly improved after applying the smoothing operation. (The LMMSE estimates after smoothing are not shown since they are essentially the same as in the ML case).

Coupled with 5-stage PIC (with noncoherent MF as the first stage), different channel estimators without smoothing are assessed and compared in Fig. 6 in terms of estimation MSE and BER performance. As expected, the ML algorithm has the worst performance. The LMMSE estimator considers the noise effect and slightly improves the estimation results, especially at low SNR (however, the BER performance is not noticeably improved). The Kalman filters take advantage of the correlation nature of the fading channel and significantly improve the estimation MSE and BER performance. The PIC is also simulated with perfect channel estimates, i.e., when $\hat{h}_k = h_k$ (which is called genie-aided PIC) to see how close the performance of proposed channel estimators is to the ideal one. It is evident from the figures that PIC with the second-order Kalman filter has the best performance, but there is still a performance penalty compared with the ideal case. The second-order Kalman filter outperforms the first-order filter at the price of higher computational complexity.

Different channel estimators combined with the smoothing filter are compared in Fig. 7. One can observe that the ML algorithm yields the smallest MSE and lowest BER after smoothing. The estimation MSE of the LMMSE algorithm is a little higher, but the BER performance is essentially the same. The ML estimator is preferable to the LMMSE estimator in that it does not need the knowledge of noise variance and the average received power of each signal path. No significant improvement is observed for the two Kalman filters after smoothing. Since the Kalman algorithm itself already takes the correlation into account, the additional smoothing does not make much difference. The ML estimator with channel smoothing appears to be the most favorable choice for estimating slow fading channels. Fig. 6.b) and Fig. 7.b) clearly show the significant improvements of the coherent PIC scheme over the noncoherent MF detection.

Based on the above experimental results, we come to the conclusions that (a) the channel estimation MSE translates into data detection BER performance, (b) the knowledge of the channel is crucial to the system performance, and (c) coherent PIC with channel estimation significantly outperforms the noncoherent MF.

Fig. 8 illustrates the convergence properties of the joint PIC with ML and LMMSE channel estimators. The variance of channel estimation error and MSE are measured before smoothing; however, the data detection is done after a 19-tap smoothing filter. The number of users $K$ is set to 12 and the stacking factor of the received vectors $D$ is set to 3 here. We examine the estimation variance of the ML estimator (which is equivalent to MSE since the ML estimator is unbiased) at each iteration and compare with the CRLB. It is shown that the ML estimator converges after three iterations. Upon convergence, we can observe very close performance agreement between ML estimation with decision feedback and the pilot-aided approach (i.e., assuming exact knowledge of the transmitted data). Only a small gap
Figure 6: Comparison of different channel estimators with the PIC without smoothing. The PIC curves represent the results at the 5th stage.

Figure 7: Comparison of different channel estimators with the PIC with smoothing. The PIC curves represent the results at the 5th stage.
is noticed at very low SNRs. The CRLB, derived in A.1, is shown to be close to the simulated estimation variance after the ML estimator converges. When the approximations expressed in (19) are used, the approximated CRLB exhibits a small discrepancy compared with the real CRLB, which is due to the fact that the approximations are based on the assumption of transmitted chip sequences from different users and different paths being orthogonal, which is too optimistic in this case. Convergence properties of the LMMSE channel estimator is examined in Fig. 8 (b). Like the ML estimator, it takes three iterations to converge, at which point the estimation MSE is close to its analytical value (22) and lower bound (23) derived in Appendix A.2.

The ML estimator and Kalman filter are compared in fast fading channels when the normalized Doppler frequency is set to \( f_d T = 0.1 \). The smoothing filter length is chosen to be 9 in this case. Fig. 9 shows that the second-order Kalman filter outperforms the ML estimator under such circumstances. The latter cannot keep track of the fast time-varying channel, and the additional filtering operation may have an undesired effect as it can destroy the details of the channel information. This is clearly shown in Fig. 9 that the ML estimator with smoothing yields higher estimation variance. However, somewhat surprisingly, the BER for the ML estimator is better with smoothing than without at high SNR. The Kalman filter, on the contrary, is capable of tracking the fast fading channels, and achieves lower estimation error as well as better BER performance compared to the ML estimator.

In Fig. 10, we examine the effect of the stacking factor \( D \) on the estimation performance. As indicated by the CRLB, it seems that the larger \( D \) is, the smaller estimation error will be. That would indeed be the case if the channel is static. However, since the real fading channel varies, the \( D \) value has to be chosen in accordance with the normalized Doppler frequency. From the plot, one can see that, for the specific channel settings in question, \( D = 4 \) appears to be the optimum value before smoothing and \( D = 2 \) or \( D = 3 \) appears to be the optimum value after smoothing. The time-varying nature of the fading channel prohibits the use of a larger stacking. Also, the dependency between stacking and smoothing, as shown by the simulation results, has to be taken into account in the selection of the stacking factor \( D \) to achieve the best channel estimation and data detection performance.

Finally, the performance of soft IC and CE is tested in an 18-user system and illustrated in Fig. 11. The improvement by using soft IC alone is not noticeable until the SNR increases to \( E_b/\eta_0 = 20 \) dB, at which point the gain by applying soft cancellation is 0.4 dB, and it is further increased to 1 dB by applying soft CE. Apparently, in order to achieve the utmost performance, the soft information should be used for both interference cancellation and channel estimation.

6 CONCLUSIONS

Multiuser detection techniques are widely proposed to combat the detrimental effects of multipath fading and MAI, which are the major impairments in CDMA communication systems. Most multiuser detectors rely on accurate channel information, which needs to be estimated in practice. In this paper, joint iterative channel estimation and interference cancellation is studied. We focus on the decision-directed approach for channel estimation, e.g.,
Figure 9: ML versus Kalman channel estimation with 5-stage PIC in fast fading channels ($f_dT = 0.1$, 9-tap smoothing filter). All the curves represent the results at the 5th stage.

Figure 10: Performance of ML channel estimation with different stacking orders.

Figure 11: Hard versus soft IC and CE. All the curves represent the results at the 5th stage.
the receiver estimates the channel parameters based on the detected data. The estimated channel coefficients are then used to regenerate the signal of each user for the purpose of interference cancellation. According to our simulation results, the performance of the proposed system is significantly better compared with the conventional system using noncoherent demodulation.

It is shown that PIC coupled with decision-directed ML estimation and an additional smoothing filter performs better than the Kalman filters in slowly fading channels. In fact, PIC and decision-directed ML channel estimation closely tracks the performance attainable by the pilot-aided channel estimation (i.e., when the transmitted data are exactly known). However, the filtering process causes a delay for channel estimation which is not desirable under some circumstances. Also the ML estimator is subject to dimensionality problem when the number of users and paths increase. This problem can be tackled by stacking the received vectors, provided the channel remains static during several symbol intervals. When this assumption is not valid, we can resort to the Kalman filter, which is suitable for tracking fast fading channels. It takes both correlation and noise into account, does not have a dimensionality problem, and does not require an additional smoothing operation. We also learned from experiments that soft IC and CE achieve better performance than to conventional IC and CE using hard decision feedback.

A APPENDIX: THEORETICAL ANALYSIS FOR CHANNEL ESTIMATION

A.1 DERIVATION OF CRLB FOR ML CHANNEL ESTIMATOR

Our main concern in this work is to estimate the complex channel gains. It is therefore of interest to establish a bound on the accuracy with which the channels can be estimated. If we restrict our attention to unbiased estimators, the natural performance measure is the error variance. The CRLB is a bound on the smallest covariance matrix that estimated. If we restrict our attention to unbiased estimators, bound on the accuracy with which the channels can be es-

- **A**
- **N**
- **r**
- **h**
- **n**
- **D**
- **L**
- **K**
- **T**
- **LL**
- **E**
- **I**
- **m**
- **l**
- **k**
- **j**
- **i**
- **s**
- **a**
- **b**
- **c**
- **d**
- **e**
- **f**
- **g**
- **h**
- **i**
- **j**
- **k**
- **l**
- **m**
- **n**
- **o**
- **p**
- **q**
- **r**
- **s**
- **t**
- **u**
- **v**
- **w**
- **x**
- **y**
- **z**

and assume the stacking factor $D$ is chosen such that the channel remains relatively static during the observation period. The vector $r$ is formed by $r = Ah + n$, and has PDF $r \sim C \mathcal{N}(Ah, N_0 I)$ where

$$A = [A^T(j) A^T(j+1) \cdots A^T(j+D-1)]^T$$

and $n = [n^T(j) n^T(j+1) \cdots n^T(j+D-1)]^T \in \mathbb{C}^{DN}$

$$= [a_{1,1} a_{1,2} \cdots a_{K,1} a_{K,2} \cdots a_{K,K}].$$

The likelihood function and log-likelihood function are

$$p(r; h) = \frac{1}{(\pi N_0)^{DN}} \exp \left[ -\frac{(r - Ah)^*(r - Ah)}{N_0} \right]$$

$$= \frac{1}{(\pi N_0)^{DN}} \exp \left[ -\frac{\|r\|^2 - h^*A^*r - r^*Ah + h^*A^*Ah}{N_0} \right]$$

$$\ln p(r; h) = -DN \ln \pi - DN \ln N_0$$

$$- \frac{\|r\|^2 - h^*A^*r - r^*Ah + h^*A^*Ah}{N_0}$$

(18)

Taking complex gradient [14] of $\ln p(r; h)$ with respect to $h$ yields

$$\frac{\partial \ln p(r; h)}{\partial h} = -\frac{1}{N_0} \frac{\partial [\|r\|^2 - h^*A^*r - r^*Ah + h^*A^*Ah]}{\partial h}$$

$$= -\frac{1}{N_0} (A^*Ah - A^*r)^*$$

The above equality holds since

$$\frac{\partial \|r\|^2}{\partial h} = 0; \quad \frac{\partial h^*A^*r}{\partial h} = 0;$$

$$\frac{\partial r^*Ah}{\partial h} = (A^*r)^*; \quad \frac{\partial h^*A^*Ah}{\partial h} = (A^*Ah)^*$$

Thus we can derive,

$$\frac{\partial \ln p(r; h)}{\partial h} = \left( \frac{\partial \ln p(r; h)}{\partial h} \right)^* = \frac{A^*r - A^*Ah}{N_0}$$

$$= \frac{A^*A}{N_0} [(A^*A)^{-1} A^*r - h] = J(h)[h] - h$$

(19)

This proves that the minimum variance unbiased (MVU) estimator of $h$ is $\hat{h} = (A^*A)^{-1} A^*r$ which is equivalent to the ML estimator described in Section 3.1. It is efficient in that it attains CRLB. The Fisher information matrix is $J(h) = A^*A/N_0$, and the covariance matrix $C_\hat{h}$ of this unbiased estimator is

$$C_\hat{h} = J^{-1}(h) = N_0(A^*A)^{-1}$$

$$\geq N_0 \text{diag}(|a_{1,1}|^2, |a_{1,2}|^2, \ldots, |a_{K,K}|^2)^{-1}$$

$$= \frac{N_0}{DN} \text{I}_{\text{tot}}$$

(19)

The inequality in (19) holds with equality when columns of $A$ are orthogonal. The CRLB for the complex
gain of the $k^{th}$ user’s $l^{th}$ path is thus the $(k-1)L+l^{th}$ diagonal element of $J^{-1}(h)$ when $L_k = L$ for all $k$, i.e.,

$$
\text{var}(h_{k,l}) = \text{diag}[C_h^{(k-1)L+l}](h_{k-1,L+l}) = \text{diag}[J^{-1}(h)]_{(k-1)L+l} \geq \frac{N_0}{DN}.
$$

(20)

### A.2 Derivation of Error Covariance for LMMSE Channel Estimator

In the following, we derive the LMMSE channel estimator and its MSE using the Bayesian Gauss-Markov Theorem. The observed data $r$ is modeled as $r = Ah + n$ where $h \in \mathbb{C}^{L_{\text{tot}}}$ is now assumed to be a random vector whose realization is to be estimated. We assume that $h$ has mean $E[h] = \mu_h = 0$ and covariance matrix $C_h = P$. The noise vector $n \in \mathbb{C}^N$ has PDF $n \sim \mathcal{CN}(0, N_0I)$ and is independent of $h$. The Bayesian MMSE estimate of $h$ that minimizes MSE averaged over all realizations of $h$ and $r$ is

$$
\hat{h} = E(h|r) = \mu_h + C_h A^*(AC_hA^* + N_0I)^{-1}(r - AP\mu_h)
= PA^*(APA^* + N_0I)^{-1}r.
$$

(21)

The performance of this estimator is measured by the error $e = h - \hat{h}$, whose mean is zero and covariance matrix is $[14]$:

$$
C_e = E(EE^T) = C_h - C_h A^*(AC_hA^* + N_0I)^{-1}AC_h
= P - PA^*(APA^* + N_0I)^{-1}AP
$$

(22)

To simplify the derivation, we assume all the paths have the same received power $P$. In this case, $P = PL$. Using the matrix inversion lemma [14] $(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$, we have

$$
(APA^* + N_0I)^{-1} = \frac{I}{N_0} - \frac{I}{N_0} A \left( \frac{I}{N_0} A + \frac{P}{N_0} \right)^{-1} A^* \frac{I}{N_0}
\leq \frac{I}{N_0} - \frac{I}{N_0} \frac{DN}{N_0} \left( \frac{I}{N_0} + \frac{P}{N_0} \right)^{-1} A^* \frac{I}{N_0} = \frac{I}{N_0} - \frac{P}{N_0(DNP + N_0)} AA^*
$$

The covariance matrix can then be computed as

$$
C_e \geq P - PA^* \left( \frac{I}{N_0} - \frac{P}{N_0(DNP + N_0)} AA^* \right) AP
\geq PI - \frac{P^2DN}{N_0} I + \frac{P^3D^2N^2}{N_0(DNP + N_0)} I
= \left[ \frac{PN_0 - P^2DN}{N_0} + \frac{P^3D^2N^2}{N_0(DNP + N_0)} \right] I
$$

The error covariance matrix is also the minimum MSE matrix [14]:

$$
\text{MSE}(h) = C_e \geq \left[ \frac{PN_0 - P^2DN}{N_0} + \frac{P^3D^2N^2}{N_0(DNP + N_0)} \right] I
$$

(23)

### References


