An Interference Suppression scheme with Joint Channel-Data Estimation *

Suhas N. Diggavi†, Boon Chong Ng ‡and A. Paulraj §

†180, Park Avenue, Information Sciences Research Center,

AT&T Shannon Laboratories, Bldg 103, Florham Park, New Jersey, NJ 07932, USA.

Email:suhas@research.att.com - Tel.: (973) 360-8492 - Fax: (973) 360-8178

‡DSO National Labs,

20 Science Park Drive, Singapore 118230, Singapore.

Email:nbooncho@dso.org.sg

§Durand, Information Systems Laboratory,

Stanford University, Stanford, CA 94305, USA.

Email:paulraj@rascals.stanford.edu

Abstract

This paper describes an adaptive space-time receiver with joint channel-data estimation (JCDE) to combat time-varying multipath channels in the presence of undesired co-channel interference (CCI). The receiver uses a colored Gaussian metric for sequence detection in order to suppress the CCI. The proposed scheme also uses the knowledge of the transmit filter for improved channel estimation to enhance performance. The algorithm is derived as a quasi-Newton scheme on a chosen cost criterion and is also locally convergent. The performance of this class of interference cancellers is examined through the pairwise error probability (PEP). Through these expressions we gain insight into the properties of the canceller. The effect of channel dynamics and identification mismatch on the PEP is also examined. To reduce implementational complexity, a hybrid delayed-decision feedback and JCDE scheme is also proposed. The performance is illustrated using numerical results in realistic transmission environments.

I Introduction

The frequency spectrum is a valuable resource in wireless communications. In order to use it efficiently, current systems have adopted a cellular structure in which the frequency spectrum is re-used by cells that are geographically separated. Frequency re-use causes co-channel interference (CCI) between cells sharing the same frequencies. If this problem is the limiting factor in performance then the system is said to be *interference limited.* Mobility of the users and the ambient environment causes dynamics in the channel characteristics. In addition delay-spread, which is a result of multiple scatterers in the radio-wave propagation environment, translates to inter-symbol interference (ISI) in a digital communication system. Hence, channel time-variation, inter-symbol interference and co-channel

*This research was supported in part by the Department of the Army, Army Research Office, under Grant No. DAAH04-95-1-0249. Additional support are provided by a DSO fellowship for B. C. Ng and an Okawa foundation fellowship for S. Diggavi.

interference constitute the three major sources of impairment in wireless channels. Of a more practical concern is receiver complexity, and this issue also has to be addressed for system designers of mobile communication receivers.

There has been extensive work on the problem of detection in the presence of ISI and channel time-variations over the past two decades. In the presence of additive white Gaussian noise and perfect channel information, the optimal minimum sequence error probability receiver is the maximum likelihood sequence estimation (MLSE) receiver using the Viterbi algorithm [1]. When the channel is time-varying (TV), adaptive equalization techniques have been proposed to track channel variations [2]. In an effort to obtain near optimal performance, an adaptive MLSE receiver has also been proposed for slow time-varying frequency-selective channels [3]. This receiver however may not perform well in fast TV channels because data are only detected after some decoding delay inherent in the Viterbi algorithm and hence the estimated channel using these detected data can be very different from the current channel. Recently, a new class of adaptive MLSE receivers has been proposed for fast TV channels which avoids the channel estimation delay problem [4, 5]. The principle of per-survivor processing (PSP) introduced in [4] provides an attractive approach to deal with joint channel and data estimation (JCDE) under unknown TV channel conditions. The PSP principle has been proposed earlier for time-invariant channels in [6] where no training preamble is required, and for maximum a-posteriori (MAP) symbol detection in [7].

Co-channel interference (CCI) presents a different and challenging problem for the mobile receiver. Interference rejection techniques have long been used by the military to suppress hostile jammers. With the possible exception of spread spectrum systems, most of these techniques rely on the use of a spatially distributed array of antenna elements at the receiver for rejecting unwanted interference. The basis of these techniques is that the interferences typically have different spatial signatures from the desired user (e.g. the angles of arrival). This motivates the need for an antenna array at the receiver for CCI mitigation in the mobile communications environment. Since the CCI may also have multipath components, temporal processing (such as equalization) may be required in addition to spatial processing. An adaptive MLSE with an MMSE space-time (ST) filter pre-processor has been proposed in [8] for slow fading channels with spatially distributed CCI. An adaptive spatial MMSE beamformer was also proposed in [9] for the IS-54 TV fading channel. A decision directed linear equalization approach based on the maximum signal-to-interference ratio was proposed in [10]. This scheme uses tentative decisions at the output of the linear equalizer and therefore could have severe error propagation. A two-stage interference cancellation approach has been proposed in [11]. The first stage is a linear equalizer suppressing interference followed by a sequence detection scheme to handle ISI. As in [8], this approach uses a decision-directed approach suitable for time-invariant channels.

More recently, interference suppression schemes based on sequence detection have been proposed [12, 13, 14, 15, 16, 17, 18]. Maximum a-posteriori (MAP) based sequence detection scheme, where both the desired signal and the interference are jointly decoded, is examined in [12, 13]. Joint detection of the desired user with the interference has also been proposed in [14, 15] where a maximum likelihood based sequence detection scheme suitable for time-invariant channels is proposed. In this paper we do not attempt to decode the undesired users and therefore do not need to know the number of interferers or their modulation schemes. In [17, 18], adaptive algorithms are used to track the channel and the interference covariance matrix. The interference covariance matrix is used in the metric calculation for sequence detection, thus suppressing CCI. In [17], a decision-directed scheme is proposed where the detected symbols are fed back after some decoding delay. A prediction scheme is then used to adapt the estimated channel impulse response (CIR). This approach could have severe error-propagation in dynamic environments. The latter approach of [18] is based on maintaining parallel adaptive estimates conditioned on candidate data sequences. As the number of possible sequences grow exponentially, only a fixed number of data sequences are retained. This Joint Channel-Data estimation with Interference Suppression (JCD-IS) scheme mitigates the effects of error propagation caused by tentative decisions and the decision delay.

In this paper we focus on the approach we proposed in [18]. We define a cost-criterion to jointly identify the channel and the interference covariance matrix. We develop a locally convergent quasi-Newton algorithm based on this cost criterion. Since the overall CIR comprises the transmit and receiver filters, it turns out that by exploiting these known filters, the total channel can be well described compactly by a structured *linear* model with fewer unknown parameters. Structured channel models have been proposed for time-invariant channels in [19, 20, 21] and for time-varying channels in [22]. In this paper we use this model to improve the joint estimation of the channel and the noise covariance matrices. After deriving the identification algorithm and describing the PSP-based detection scheme, we analyze its performance by using the pairwise error probability (PEP). We derive the Chernoff upper bound when we have perfect channel state information (CSI) and noise covariance. This expression shows us that the performance of the interference canceller is similar to that of a space-time MLSE detector in a reduced dimensional space. We also give expressions for the PEP under channel mismatch and the impact of channel dynamics on performance is examined. To reduce complexity of the receiver, we incorporate a delayed-decision feedback sequence estimation (DDFSE) [23] scheme into the JCD-IS receiver. This effectively reduces the number of states in the trellis. Each state in this reduced trellis has an associated partial state which together gives the full state information. The impact of non-synchronous CCI on the performance is examined through numerical simulations.

This paper is organized as follows. In Section II, we introduce the notation and the data model used. In Section III, we describe the receiver structure and the proposed detection scheme. The PEP expressions are derived in Section IV. Some practical issues such as complexity and abrupt CCI variations are addressed in Section V. Numerical results are presented in Section VI, illustrating performance of these schemes in realistic channel environments. We conclude with a short discussion in Section VII.

II A Discrete Data Model

Consider a base station with M antenna elements. Let $c^i(t,\tau)$ be the impulse response of the propagation channel from the output of the transmit filter of the mobile to the ith antenna at time t due to an input unit impulse applied at time $t-\tau$. Then the impulse response is

$$h^i(t,\tau) = \int_0^\infty c^i(t,\beta)\, g(\tau-\beta) d\beta \quad i = 1, \cdots, M. \,.$$

The signal received at the ith antenna can be written as

$$x^i(t) = \sum_n s_n \int_0^\infty c^i(t,\beta) g(t-\beta-nT) d\beta + z^i(t) \quad i = 1, \cdots M. \tag{1}$$

where $g(t)$ is the transmit filter, s_n is the user data, T is the symbol period and $z^i(t)$ is the effect of CCI and the receiver noise.

The receiver consists of an anti-aliasing filter and a sampler of rate Q/T. Let us assume that the input is bandlimited to W_I and the channel time-variation bandwidth (also called the system bandwidth W_s [24]) is finite. Then we have Nyquist sampling for rate larger than $2(W_I + W_s)$. Thus after collecting sufficient statistics by Nyquist sampling we have an equivalent discrete-time model. Furthermore, we can approximate the equivalent discrete-time infinite impulse response channel by a finite impulse response (FIR) filter with an error that can be made arbitrarily small by choosing an appropriate length (L) equivalent FIR channel. Suppose that the $c^i(t,\tau)$ is causal and of finite duration, that is $c^i(t,\tau) = 0 \;\forall\, \tau < 0$ and $\tau > T_c$. It can be shown as in [19] that by approximating the integral in (1) by a finite Riemann sum of ν terms, the $Q \times 1$ over-sampled output vector at the kth epoch can be written as

$$\begin{aligned} \mathbf{x}_k^i &\triangleq \left[x^i(kT+t_o), \cdots, x^i(kT+(Q-1)T/Q+t_o)\right]^T \\ &= \widetilde{\mathbf{G}}^T(\mathbf{I}_L \otimes \mathbf{c}_k^i)\,\mathbf{s}_k + \mathbf{z}_k^i \qquad i = 1, \cdots, M \end{aligned} \tag{2}$$

where t_0 is the timing offset, $\mathbf{I}_L$ denotes the $L \times L$ identity matrix, $\mathbf{s}_k = [s_k, s_{k-1}, \cdots, s_{k-L+1}]^T$, $\mathbf{c}_k^i$ denotes a $\nu \times 1$ unknown TV parameter vector, $\otimes$ denotes the Kronecker matrix product and $\widetilde{\mathbf{G}}$ is a matrix constructed from samples of the transmit pulse shaping filter and is given by

$$\widetilde{\mathbf{G}} = \begin{bmatrix} \mathbf{g}_{00} & \cdots & \mathbf{g}_{0(Q-1)} \\ \vdots & \vdots & \vdots \\ \mathbf{g}_{(L-1)0} & \cdots & \mathbf{g}_{(L-1)(Q-1)} \end{bmatrix}_{\nu L \times Q},$$

Here $\mathbf{g}_{ij}$ is given by

$$\mathbf{g}_{ij} = [g(t_o+iT+jT/Q), g(t_o+iT+jT/Q-\Delta_c), \cdots, g(t_o+iT+jT/Q-(\nu-1)\Delta_c)]^T$$

where $\Delta_c = T_c/\nu$. In (2) we have assumed that $c^i(t,\tau)$ is constant within a symbol interval. This can be easily removed at the cost of having more parameters in the vector $\mathbf{c}_k^i$.

By stacking all the sampled antenna outputs $\mathbf{x}_k = [\mathbf{x}_k^{1T}, \ldots, \mathbf{x}_k^{MT}]^T$, we get

$$\mathbf{x}_k = \mathbf{H}_k\,\mathbf{s}_k + \mathbf{z}_k \tag{3}$$

where

$$\mathbf{H}_k = \begin{bmatrix} \widetilde{\mathbf{G}}^T\left(\mathbf{I}_L \otimes \mathbf{c}_k^1\right) \\ \vdots \\ \widetilde{\mathbf{G}}^T\left(\mathbf{I}_L \otimes \mathbf{c}_k^M\right) \end{bmatrix}_{MQ \times L} \tag{4}$$

Let the $Q \times \nu$ matrix $\overline{\mathbf{G}}_i$ $(i = 0, 1, \cdots, L-1)$ of pulse shape samples be defined as

$$\overline{\mathbf{G}}_i \triangleq \begin{bmatrix} \mathbf{g}_{i0}^T \\ \mathbf{g}_{i1}^T \\ \vdots \\ \mathbf{g}_{i(Q-1)}^T \end{bmatrix}_{Q \times \nu}$$

The vector channel $\mathbf{h}_k$ defined as $\mathbf{h}_k = \mathrm{vec}(\mathbf{H}_k)$ can be written as

$$\mathbf{h}_k = \overline{\mathbf{G}}\mathbf{c}_k \tag{5}$$

where $\mathbf{c}_k = [(\mathbf{c}_k^1)^T, (\mathbf{c}_k^2)^T, \cdots, (\mathbf{c}_k^M)^T]^T$ and

$$\overline{\mathbf{G}} = \begin{bmatrix} \mathbf{I}_M \otimes \overline{\mathbf{G}}_0 \\ \mathbf{I}_M \otimes \overline{\mathbf{G}}_1 \\ \vdots \\ \mathbf{I}_M \otimes \overline{\mathbf{G}}_{L-1} \end{bmatrix}_{MQL \times M\nu}$$

With this definition we can also write the data model as

$$\mathbf{x}_k = \mathbf{B}_k \mathbf{c}_k + \mathbf{z}_k \tag{6}$$

where $\mathbf{B}_k = (\mathbf{s}_k^T \otimes \mathbf{I}_{MQ})\overline{\mathbf{G}}$.

The goal of the proposed receiver is to estimate jointly the data $\mathbf{s}_k$ and the channel structure vector $\mathbf{c}_k$ in the presence of $\mathbf{z}_k$ (noise and CCI).

III Interference suppression scheme

The goal of an interference suppression scheme is to identify the desired user's channel and data while suppressing the undesired interference. In general this is a hard proposition and we present one possible approach. We begin the discussion by assuming that we know the desired signal's data sequence. Using this knowledge we propose an adaptive algorithm to jointly estimate the CIR of the desired signal and the noise covariance matrix. In subsection A we start with a heuristic argument justifying the algorithm. In subsection B we derive the algorithm as a locally convergent quasi-Newton scheme for a chosen cost criterion. Using the proposed identification scheme for known data sequences, we propose a tracking mode algorithm based on the per-survivor principle (PSP). This scheme along with an improved channel estimation procedure is described in subsection C.

A Heuristic Argument

Given the model in (6), and the noise covariance matrix $\mathbf{R} <_{\mathbf{z}}$, the natural criterion for channel identification is:

$$\mathbf{c}_{opt} = \arg\min_{\mathbf{c}} \sum_k (\mathbf{x}_k - \mathbf{B}_k \mathbf{c})^H \mathbf{R}_{\mathbf{z}}^{-1} (\mathbf{x}_k - \mathbf{B}_k \mathbf{c}) \tag{7}$$

This is easily implemented recursively by a weighted RLS algorithm (WRLS) [25], which could potentially be used in non-stationary environments. However, the noise covariance matrix is typically unknown. Therefore a reasonable choice of $\mathbf{R}_{\mathbf{z}}$ in implementing the WRLS algorithm would be to choose,

$$\mathbf{R}_{\mathbf{z},k} = \frac{1}{k} \sum_{i=1}^{k} (\mathbf{x}_i - \mathbf{B}_i \hat{\mathbf{c}}_k)(\mathbf{x}_i - \mathbf{B}_i \hat{\mathbf{c}}_k)^H \tag{8}$$

However, the overall algorithm would not be recursive as (8) cannot be computed recursively. In order to develop a recursive algorithm a further approximation could be made by replacing $\hat{\mathbf{c}}_k$ in (8) by $\hat{\mathbf{c}}_i$. Here $\hat{\mathbf{c}}_k$ and $\hat{\mathbf{c}}_i$ respectively are the estimates of the channel after k and i samples of data have been collected, with $i \leq k$. This would be a good approximation if the parameter estimates $\hat{\mathbf{c}}_i$ are close to $\hat{\mathbf{c}}_k$ (*i.e.* close to convergence). A weighted mean where more weight is put on the recent estimates,

$$\mathbf{R}_{\mathbf{z},k} = \frac{1}{k} \sum_{i=1}^{k} w^{k-i} (\mathbf{x}_i - \mathbf{B}_i \hat{\mathbf{c}}_i)(\mathbf{x}_i - \mathbf{B}_i \hat{\mathbf{c}}_i)^H \tag{9}$$

where w is a forgetting factor, is desirable in a time-varying environment. As $\mathbf{R}_{\mathbf{z},k}^{-1/2}$, is needed for the WRLS algorithm, we can use a recursive square-root algorithm [25] to update (9). This allows us to construct the identification algorithm using decoupled recursions for $\hat{\mathbf{c}}_k$ and $\mathbf{R}_{\mathbf{Z},k}$. The former is a weighted RLS recursion and the latter can be computed using a recursive square-root algorithm [25]. Therefore the identification algorithm could be summarized

Given $\mathbf{Q}_{k-1}, \mathbf{x}_k, \mathbf{H}_{k-1}$, $\mathbf{P}_{k-1}$ and $\hat{\mathbf{s}}_k$,
0. $\bar{\mathbf{B}}_k = \mathbf{Q}_{k-1}\mathbf{B}_k$
1. $\mathbf{K}_k = \lambda^{-1}\mathbf{P}_{k-1}\bar{\mathbf{B}}_k^H[\mathbf{I}+\lambda^{-1}\bar{\mathbf{B}}_k\mathbf{P}_{k-1}\bar{\mathbf{B}}_k^H]^{-1}$
2. $\boldsymbol{\psi}_k = \mathbf{Q}_{k-1}(\mathbf{x}_k - \mathbf{B}_k\hat{\mathbf{c}}_{k-1})$
3. $\hat{\mathbf{c}}_k = \hat{\mathbf{c}}_{k-1} + \mathbf{K}_k\boldsymbol{\psi}_k$
4. $\mathbf{P}_k = \lambda^{-1}\mathbf{P}_{k-1} - \lambda^{-1}\mathbf{K}_k\bar{\mathbf{B}}_k\mathbf{P}_{k-1}$
5. $\mathbf{y}_k = \mathbf{x}_k - \mathbf{H}_k\hat{\mathbf{s}}_k$.
6. $\gamma_k = \sqrt{w + \|\mathbf{Q}_{k-1}\mathbf{y}_k\|^2}$.
7. $\mathbf{p}_k = \mathbf{Q}_{k-1}\mathbf{y}_k/\gamma_k$.
8. $\beta_k = \left(1-\sqrt{1-\|\mathbf{p}_k\|^2}\right)/\|\mathbf{p}_k\|^2$.
9. $\mathbf{Q}_k = \frac{1}{w}\left(\mathbf{I}-\beta_k\mathbf{p}_k\mathbf{p}_k^H\right)\mathbf{Q}_{k-1}$.

Table 1: The Identification Algorithm

as in Table 1. Here we have denoted ([25], Chapter 13) the gain matrix by $\mathbf{K}_k$, the inverse of the correlation matrix by $\mathbf{P}_k$, and the inverse square root of the noise covariance (weighting) matrix by $\mathbf{Q}_k = \mathbf{R}_{\mathbf{Z},k}^{-1/2}$. The steps 1–4 are just the weighted least squares algorithm with the weighting matrix given by $\mathbf{R}_{\mathbf{Z},k}^{-1}$. The steps 5–9 is the update step for $\mathbf{Q}_k$ using a square-root algorithm. In step 5 the computation of $\mathbf{H}_k$ is done by using the estimate $\hat{\mathbf{c}}_{k-1}$ in (4).

There are two things that are quite unclear in the foregoing discussion. First, by introducing the approximation in (9), it is not evident what the criterion the algorithm in Table 1 is minimizing. Second, it is not obvious that such a scheme would converge even locally. These two issues are studied in the next subsection.

B The Cost Criterion

Consider the problem,

$$\min_{\mathbf{c},\mathbf{R}} J_k(\mathbf{c},\mathbf{R}) \triangleq \frac{1}{k}\sum_{i=1}^{k}(\mathbf{x}_i-\mathbf{B}_i\mathbf{c})^H\mathbf{R}^{-1}(\mathbf{x}_i-\mathbf{B}_i\mathbf{c}) + \text{logdet}\mathbf{R} \tag{10}$$

Note that this criterion is equivalent to the maximum likelihood criterion if the noise $\mathbf{z}_i \sim \mathbb{C}\mathcal{N}(0,\mathbf{R})$ is i.i.d. Let us define an operator $\hat{\mathbb{E}}_k[\cdot]$ as,

$$\hat{\mathbb{E}}_k[\mathbf{U}] = \frac{1}{k}\sum_{i=1}^{k}\mathbf{U}_i \tag{11}$$

Claim III.1 *The stationary points of $J_k(\mathbf{c},\mathbf{R})$ given in (10) are represented as,*

$$\begin{aligned}\mathbf{c} &= \hat{\mathbb{E}}_k[\mathbf{B}^H\mathbf{R}^{-1}\mathbf{B}]^{-1}\hat{\mathbb{E}}_k[\mathbf{B}^H\mathbf{R}^{-1}\mathbf{x}] \\ \mathbf{R} &= \hat{\mathbb{E}}_k[(\mathbf{x}-\mathbf{B}\mathbf{c})(\mathbf{x}-\mathbf{B}\mathbf{c})^H]\end{aligned} \tag{12}$$

Proof: It can easily be verified that:

$$\begin{aligned}\frac{\partial J_k(\mathbf{c},\mathbf{R})}{\partial\mathbf{c}} &= -\hat{\mathbb{E}}_k[\mathbf{B}^H\mathbf{R}^{-1}(\mathbf{x}-\mathbf{B}\mathbf{c})] \\ \frac{\partial J_k(\mathbf{c},\mathbf{R})}{\partial\mathbf{R}} &= -\mathbf{R}^{-1}\hat{\mathbb{E}}_k[(\mathbf{x}-\mathbf{B}\mathbf{c})(\mathbf{x}-\mathbf{B}\mathbf{c})^H]\mathbf{R}^{-1} + \mathbf{R}^{-1}\end{aligned} \tag{13}$$

For the stationary points, $\frac{\partial J_k(\mathbf{c},\mathbf{R})}{\partial \mathbf{c}} = 0$ and $\frac{\partial J_k(\mathbf{c},\mathbf{R})}{\partial \mathbf{R}} = 0$, and hence we obtain (12) ■

This result indicates that we might be optimizing the criterion (10) in the algorithm described in Table 1. The algorithm described in subsection III.A belongs to the general class of recursive algorithms studied extensively in literature [26]. The identification algorithm described here is a special case of the recursive prediction error method (RPEM) studied in [26]. Hence the question about convergence of this algorithm can be answered by using Theorem 4.4 of [26] where it is shown that a general class of recursive identification schemes are convergent. The algorithm described in Table 1 belongs to this class and hence is convergent. Note that the criterion described in (10) is the ML criterion for identification in temporally white Gaussian noise with spatial covariance matrix $\mathbf{R}$. This criterion has also been applied to spatial regressor estimation in colored noise in [27]. There the noise covariance was parametrized and jointly identified with the regressors. Our identification scheme does not assume any structure for the noise covariance matrix. Given the criterion in (10) it is natural to expect that we can explicitly derive the identification algorithm as a quasi-Newton scheme. To the best of our knowledge, this has not been reported in literature, and it is instructive to make this connection. We will present the derivation for the real case and the extension to the complex case is not difficult.

Let us define the vector of unknown parameters as $\boldsymbol{\psi} = [\mathbf{c}^T, \text{vec}(\mathbf{R})^T]^T$. By using (13) we can easily write,

$$\frac{\partial J_k(\mathbf{c},\mathbf{R})}{\partial \boldsymbol{\psi}} = \begin{bmatrix} -\hat{\mathbb{E}}_k[\mathbf{B}^T\mathbf{R}^{-1}(\mathbf{x}-\mathbf{B}\mathbf{c})] \\ \text{vec}\left(-\mathbf{R}^{-1}\hat{\mathbb{E}}_k[(\mathbf{x}-\mathbf{B}\mathbf{c})(\mathbf{x}-\mathbf{B}\mathbf{c})^T]\mathbf{R}^{-1}+\mathbf{R}^{-1}\right) \end{bmatrix} \tag{14}$$

To derive the quasi-Newton algorithm we use:

$$\hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_{k-1} - [\frac{\partial^2 J_k(\hat{\boldsymbol{\psi}}_{k-1})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}^T}]^{-1}\frac{\partial J_k(\hat{\boldsymbol{\psi}}_{k-1})}{\partial \boldsymbol{\psi}} \tag{15}$$

We can write,

$$\mathbf{V} \triangleq \frac{\partial^2 J_k(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}^T} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}^T & \mathbf{V}_{22} \end{bmatrix} \tag{16}$$

where $\mathbf{V}_{11} = \frac{\partial^2 J_k(\mathbf{c},\mathbf{R})}{\partial \mathbf{c}\partial \mathbf{c}^T}$, $\mathbf{V}_{12} = \frac{\partial^2 J_k(\mathbf{c},\mathbf{R})}{\partial \mathbf{c}\partial \text{vec}(\mathbf{R})^T}$, and $\mathbf{V}_{22} = \frac{\partial^2 J_k(\mathbf{c},\mathbf{R})}{\partial \text{vec}(\mathbf{R})\partial \text{vec}(\mathbf{R})^T}$. It can easily be verified that

$$\mathbf{V}_{11} = \hat{\mathbb{E}}_k[\mathbf{B}^T\mathbf{R}^{-1}\mathbf{B}] \tag{17}$$

For the case when we are close to the optimal solution, $\mathbf{V}_{12} = \frac{\partial}{\partial \text{vec}(\mathbf{R})^T}(\hat{\mathbb{E}}_k[\mathbf{B}^T\mathbf{R}^{-1}(\mathbf{B}\mathbf{c}-\mathbf{x})]) \approx 0$. This can be proved more formally by explicit calculation and using the assumption that we are close to the optimal solution. Hence we find that $\mathbf{V}$ is a block diagonal matrix.

Let us define $\boldsymbol{\alpha} = \mathbf{x} - \mathbf{B}\mathbf{c}$.

Claim III.2

$$\mathbf{V}_{22}|_{\hat{\mathbb{E}}_k[\boldsymbol{\alpha}\boldsymbol{\alpha}^T]=\mathbf{R}} = \mathbf{R}^{-1}\otimes\mathbf{R}^{-1} \tag{18}$$

Proof: See Appendix A.

Using results (15), (17) Claim III.2 and (14) we obtain decoupled¡ recursions for $\hat{\mathbf{c}}_k$ and $\mathbf{R}_k$. For notational convenience we interchangably use $\mathbf{R}_k$ and $\mathbf{R}_{\mathbf{Z},k}$. The recursion for $\hat{\mathbf{c}}_k$ is,

$$\hat{\mathbf{c}}_k = \hat{\mathbf{c}}_{k-1} + [\sum_{i=1}^{k}\mathbf{B}_i^T\mathbf{R}_{k-1}^{-1}\mathbf{B}_i]^{-1}\mathbf{B}_k^T\mathbf{R}_{k-1}^{-1}(\mathbf{x}_k - \mathbf{B}_k\hat{\mathbf{c}}_{k-1}) \tag{19}$$

This is obtained by noticing that close to convergence we have $\hat{\mathbb{E}}_{k-1}[\mathbf{B}^T\mathbf{R}^{-1}(\mathbf{x}-\mathbf{B}\mathbf{c})] \approx 0$. This is identical to the steps in [26] (Chapter 3) to derive the WRLS algorithm. To get a recursive form for calculating $\mathbf{V}_{11}$ we need to approximate $\mathbf{R}_{k-1}$ in (19) by $\mathbf{R}_{i-1}$, which is a good approximation close to convergence. Here $\mathbf{R}_{k-1}$ and $\mathbf{R}_{i-1}$ are respectively the estimates of $\mathbf{R}_{\mathbf{Z}}$ after $k-1$ and $i-1$ samples of data have been collected with $i \leq k$. Using this modification the recursion in (19) is exactly the update step used in Table 1. To update $\mathbf{R}_k$ we have,

$$\text{vec}(\mathbf{R}_k) = \text{vec}(\mathbf{R}_{k-1}) + (\mathbf{R}_{k-1}^{-1} \otimes \mathbf{R}_{k-1}^{-1})^{-1}\text{vec}(\mathbf{R}_{k-1}^{-1}\hat{\mathbb{E}}_k[(\mathbf{x}-\mathbf{B}\mathbf{c})(\mathbf{x}-\mathbf{B}\mathbf{c})^T]\mathbf{R}_{k-1}^{-1} - \mathbf{R}_{k-1}^{-1}) \tag{20}$$

We use the fact that $\hat{\mathbb{E}}_k[(\mathbf{x}-\mathbf{B}\mathbf{c})(\mathbf{x}-\mathbf{B}\mathbf{c})^T] = \frac{(k-1)}{k}\hat{\mathbb{E}}_{k-1}[(\mathbf{x}-\mathbf{B}\mathbf{c})(\mathbf{x}-\mathbf{B}\mathbf{c})^T] + \frac{1}{k}(\mathbf{x}_k - \mathbf{B}_k\hat{\mathbf{c}}_{k-1})(\mathbf{x}_k - \mathbf{B}_k\hat{\mathbf{c}}_{k-1})^T$. Also, close to convergence $\hat{\mathbb{E}}_{k-1}[(\mathbf{x}-\mathbf{B}\mathbf{c})(\mathbf{x}-\mathbf{B}\mathbf{c})^T] \approx \mathbf{R}_{k-1}$. Using this we obtain

$$\text{vec}(\mathbf{R}_k) = \frac{(k-1)}{k}\text{vec}(\mathbf{R}_{k-1}) + \frac{1}{k}\text{vec}\left((\mathbf{x}_k - \mathbf{B}_k\hat{\mathbf{c}}_{k-1})(\mathbf{x}_k - \mathbf{B}_k\hat{\mathbf{c}}_{k-1})^T\right) \tag{21}$$

To obtain the last term we used (D.1) and (D.2). This is the desired recursion for $\mathbf{R}_k$ as described in (9). Thus from the above analysis we have derived the algorithm described in Table 1 as a quasi-Newton recursive algorithm for criterion (10).

C The Detection Scheme

In the discussions above we had assumed that we had access to the correct data sequence. However, this is not true except during the training period. Therefore, in the tracking mode, this problem is typically handled by decision-directed adaptation. However, as this could lead to severe error propagation, the principle of per-survivor processing [7, 6, 4] could be used to mitigate the problem. Given the CIR and the noise covariance matrix, the optimal detection scheme is a maximum likelihood (ML) scheme. In the presence of undesired CCI, it is impractical to estimate the probability distribution of the noise and hence ML detection is difficult. Recent information-theoretic results in mismatched detection [28], have shown that a Gaussian capacity can be achieved by using a Gaussian codebook and a Gaussian decoding scheme. Thus even though the detection is not ML (and hence mismatched), with powerful enough Gaussian coding schemes, we can still achieve Gaussian rates. This result motivates us to use a Gaussian decoding metric based on the noise-covariance for detection. The information theoretic results [28] indicate that we could asymptotically achieve performance of the equivalent Gaussian channel.

If we use a Gaussian decoding scheme for a noise which is assumed to be white with spatial covariance matrix $\mathbf{R}_{\mathbf{Z}}$, we have a branch metric,

$$\left\|\mathbf{R}_{\mathbf{Z}}^{-1/2}(\mathbf{x}_k - \mathbf{H}_k\hat{\mathbf{s}}_k)\right\|^2 = \left\|\mathbf{R}_{\mathbf{Z}}^{-1/2}(\mathbf{x}_k - \hat{\mathbf{B}}_k\mathbf{c}_k)\right\|^2 .$$

If the noise is temporally colored (as would typically be the case with CCI) one would ideally require a discrete-time noise whitening filter for ML decoding. In practical terms, this increases the number of parameters to be estimated and tracked and also the effective channel length. Since the main structure is typically in the spatial covariance, we adopt the above decoding metric that does not take the time-correlation of the CCI into account. Hence we only track the $MQ \times MQ$ coloring matrix. However, the temporal correlation of the CCI could easily be incorporated at the cost of higher complexity. As we only have access to estimates $\hat{\mathbf{R}}_{\mathbf{Z},k}$ of the noise covariance matrix and the channel estimate $\hat{\mathbf{H}}_k$ we propose the following decoding branch metric

$$\text{BM}_k = \left\|\mathbf{R}_{\mathbf{Z},k}^{-1/2}(\mathbf{x}_k - \hat{\mathbf{H}}_k\hat{\mathbf{s}}_k)\right\|^2 = \left\|\mathbf{R}_{\mathbf{Z},k}^{-1/2}(\mathbf{x}_k - \hat{\mathbf{B}}_k\hat{\mathbf{c}}_k)\right\|^2 . \tag{22}$$

Here the estimates $\mathbf{R}_{\mathbf{z},k}^{-1/2}$ and $\hat{\mathbf{c}}_k$ are obtained by using the algorithm summarized in Table 1, and we incorporate the knowledge of the transmit pulse shape into the WRLS scheme to provide improved channel estimates. If $\mathbf{R_Z} = \mathbf{I}$ is used, then the detection scheme would be a minimum Euclidean distance decoding (MEDD) scheme.

Let $\mathcal{I}$ denote the admissible set of state indices that can lead to state i, $\mathbf{s}^{ji}$ denote the $L \times 1$ symbol vector associated with state j and the transition to state i. Let the functionals $f_c(\cdot)$ and $f_Q(\cdot)$ denote the adaptive algorithms (Table 1) for updating the structured parameter vector and the coloring matrix $\mathbf{Q}$ respectively. Then the algorithm for the receiver is as follows:
For each i (state index),

1. Select previous state index corresponding to minimum path metric
   $J = \arg\min_{j \in \mathcal{I}} \left\{ C_{k-1}^{(j)} + \left\| \mathbf{Q}_{k-1}^{(j)} (\mathbf{x}_k - \mathbf{H}_{k-1}^{(j)} \mathbf{s}^{ji}) \right\|^2 \right\}$,
2. Update survivor path metric for ith state
   $C_k^{(i)} = C_{k-1}^{(J)} + \left\| \mathbf{Q}_{k-1}^{(J)} (\mathbf{x}_k - \mathbf{H}_{k-1}^{(J)} \mathbf{s}^{Ji}) \right\|^2$,
3. Extend survivor path for the ith state
   $SV_k^{(i)} = \mathbf{s}^{Ji} \bigcup SV_{k-1}^{(J)}$,
4. Update $\mathbf{Q}_k^{(i)}$
   $\mathbf{Q}_k^{(i)} = f_Q \left(\mathbf{Q}_{k-1}^{(J)}, \mathbf{x}_k, SV_k^{(i)}, \mathbf{H}_{k-1}^{(J)} \right)$,
5. Update structured channel matrix
   $\mathbf{c}_k^{(i)} = f_c \left(\{\mathbf{x}_n\}_{n=1}^{k}, \mathbf{c}_{k-1}^{(J)}, SV_k^{(i)}, \mathbf{Q}_k^{(i)} \right)$,
   $\mathbf{H}_k^{(i)}$ is constructed from (4) .

IV Analysis

In this section we examine the pairwise error probability (PEP) of sequence detection in the presence of a spatially colored interference. We examine the impact of channel dynamics on the PEP. To do this, we calculate the approximate PEP in the presence of channel identification errors (channel mismatch). The Chernoff bound for the PEP is derived in subsection A. In subsection B we present the computation of the PEP under mismatched conditions.

A Chernoff bound analysis

The PEP is a very useful tool in analyzing the performance of a sequence detection scheme [29]. For the Chernoff bound we analyze the problem when the interference is temporally white. This allows us to gain insight into the problem by examining the analytical results. The probability that the correct sequence $\mathbf{s}^{(0)}$ is mistaken for the incorrect one $\mathbf{s}^{(1)}$ when we have perfect CSI is:

$$\begin{aligned}\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}|\boldsymbol{\xi}) &= \Pr\Bigg[\sum_{k=k_1}^{k_1+L_e-1}(\mathbf{x}_k - \widetilde{\mathbf{B}}_k^{(0)}\mathbf{c})^H\mathbf{R}_{\mathbf{z}}^{-1}(\mathbf{x}_k - \widetilde{\mathbf{B}}_k^{(0)}\mathbf{c}) \\ &\qquad \geq \sum_{k=k_1}^{k_1+L_e-1}(\mathbf{x}_k - \widetilde{\mathbf{B}}_k^{(1)}\mathbf{c})^H\mathbf{R}_{\mathbf{z}}^{-1}(\mathbf{x}_k - \widetilde{\mathbf{B}}_k^{(1)}\mathbf{c})\Bigg] \\ &\overset{(a)}{\leq} \mathbb{E}[\exp(\gamma\sum_{k=k_1}^{k_1+L_e-1}\mathbf{z}_k^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{z}_k - (\mathbf{E}_k\mathbf{c}+\mathbf{z}_k)^H\mathbf{R}_{\mathbf{z}}^{-1}(\mathbf{E}_k\mathbf{c}+\mathbf{z}_k))]\end{aligned} \tag{23}$$

where (a) follows from the Chernoff bound, $\boldsymbol{\xi}$ is the interference symbol sequence, $\mathbf{E}_k = \widetilde{\mathbf{B}}_k^{(0)} - \widetilde{\mathbf{B}}_k^{(1)}$ and $\widetilde{\mathbf{B}}_i^{(l)} = (\mathbf{s}_i^{(l)})^T \otimes \mathbf{I}_{MQ}$. Here we have assumed that the CIR and the noise covariance matrix are known. Also $\mathbf{c}$ (the physical channel) and the noise covariance ($\mathbf{R}_{\mathbf{Z}}$) do not vary over the error event. Moreover as we have conditioned on the interference symbols, the noise is Gaussian. Note that if we assume that the U interferers are narrowband then $\mathbf{z}_k = \sum_{u=1}^{U}\mathbf{h}^{(u)}\xi_k^{(u)} + \tilde{\mathbf{z}}_k$, where $\tilde{\mathbf{z}}_k$ is the AWGN and $\mathbf{h}^{(u)}$ is the (flat) fading vector channel of the u^{th} interferer. Hence we have,

$$\mathbb{E}[\mathbf{z}_k\mathbf{z}_k^H|\boldsymbol{\xi}] = \sum_{u=1}^{U}\mathbf{R}_{\mathbf{h}}^{(u)}|\xi_k^{(u)}|^2 + \sigma^2\mathbf{I} \tag{24}$$

Here we have defined $\mathbf{R}_{\mathbf{h}}^{(u)} = \mathbb{E}\mathbf{h}^{(u)}\mathbf{h}^{(u)H}$. If we assume that the interferers have a constant modulus modulation, the noise covariance matrix conditioned on the interference symbols is the same as the unconditioned one. Using this assumption, we can now write the PEP as,

$$\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}) \leq \mathbb{E}_{\mathbf{c}}[e^{-\gamma(1-\gamma)\mathbf{c}^H(\sum_k \mathbf{E}_k^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{E}_k)\mathbf{c}}] \tag{25}$$

where we have averaged (23) over $\mathbf{z}_k$. By optimizing the Chernoff parameter γ and averaging over $\mathbf{c} \sim \mathbb{C}\mathcal{N}(0, \mathbf{R}^{(\mathbf{C})})$, we obtain,

$$\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}) \leq \frac{1}{|\mathbf{I} + \frac{1}{4}\mathbf{R}^{(\mathbf{C})1/2}\sum_k \mathbf{E}_k^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{E}_k\mathbf{R}^{(\mathbf{C})H/2}|} \tag{26}$$

To gain insight into (26), we assume that each tap in the CIR $\mathbf{c}$ fades independently and identically (WSSUS model). This implies that $\mathbf{R}^{(\mathbf{C})} = \mathbb{E}\mathbf{c}\mathbf{c}^H$ can be written as,

$$\mathbf{R}^{(\mathbf{C})} = \mathbf{I}_L \otimes \mathbf{R}_c \tag{27}$$

where $\mathbf{R}_c$ is a $(MQ) \times (MQ)$ matrix [1] which determines the covariance of each tap in the vector channel. Note that we can write,

$$\mathbf{E}_k^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{E}_k = \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H \otimes \mathbf{R}_{\mathbf{z}}^{-1} \tag{28}$$

where $\boldsymbol{\epsilon}_k = [e_k, \ldots, e_{k-L+1}]^T$ and $e_k = \mathbf{s}_k^{(0)} - \mathbf{s}_k^{(1)}$. Using (27) and (28) in (26) we get,

$$\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}) \leq \frac{1}{|\mathbf{I} + \frac{1}{4}(\sum_k \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H) \otimes (\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{R}_c)|} \tag{29}$$

[1] Note that this is a modeling assumption to simplify the interpretation of the PEP expression. Also note that $\mathbf{R}^{(\mathbf{C})}$ is *not* the covariance of $\mathbf{c}$ which is the channel approximation in the structured channel model. Also, this reflects the correlation among antenna elements. The independence assumption in (27) is among time taps and not the antennas, *i.e.* it is not a diversity array.

where we have used (D.3). Using the result (D.4) we can rewrite (29) as,

$$\mathrm{PEP}(\mathbf{s}^{(0)} \rightarrow \mathbf{s}^{(1)}) \leq \prod_i \prod_k \frac{1}{1+\frac{1}{4}\lambda_k(\sum_k \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H)\lambda_i(\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{R}_c)} \tag{30}$$

where $\lambda_i(\mathbf{A})$ is the eigenvalue of matrix $\mathbf{A}$. This expression yields some insight into the performance of the interference cancellation scheme. Suppose that the interference lies in a D_{int} dimensional subspace of $\mathbb{C}^{MQ}$. Then $\mathbf{R}_{\mathbf{z}}$ would have D_{int} dominant eigenvalues and hence $\mathbf{R}_{\mathbf{z}}^{-1}$ would have $MQ - D_{int}$ dominant eigenvalues. The expression in (30) is mostly determined by eigenvalues of $\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{R}_c$ which are not close to zero. Hence the $\mathbf{R}_{\mathbf{z}}^{-1}$ operation gives little weight to the interference subspace and therefore suppresses it. Let the dominant subspace of the signal be of dimension D_s. If the dimension of the intersection between the dominant interference and signal subspaces is given by D_{com}, then the PEP will behave as if it had $(D_s - D_{com})$ diversity elements. If the signal occupied the entire $\mathbb{C}^{MQ}$ space then the PEP would behave as if it had $(MQ - D_{int})$ diversity elements. This kind of insight has been obtained in [30, 31] when the interference channels are known and perfectly nulled. Note that in (30) the eigenvalues of $\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{R}_c$ are the generalized eigenvalues of the matrix pencil $(\mathbf{R}_c, \mathbf{R}_{\mathbf{z}})$.

B Pairwise Error Probability

In this section, we derive the PEP for the JCD-IS receiver when we have channel estimation errors. The PEP for fading channels with perfect channel side-information has been examined in [32]. The performance of coding schemes in ISI-free scalar fading channels for particular channel estimation schemes has been studied in [33, 34]. We first derive the PEP for an adaptive channel estimation algorithm and PSP in the presence of CCI. We later derive an approximation for this case which is less computationally intensive to evaluate.

Conditioned on the data sequence of the CCI $\boldsymbol{\xi}$, the PEP is

$$\mathrm{PEP}(\mathbf{s}^{(0)} \rightarrow \mathbf{s}^{(1)}|\boldsymbol{\xi}) = P\left(\sum_{i=1}^{N}\left\|\mathbf{x}_i - \hat{\mathbf{H}}_i^{(0)}\mathbf{s}_i^{(0)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} > \sum_{i=1}^{N}\left\|\mathbf{x}_i - \hat{\mathbf{H}}_i^{(1)}\mathbf{s}_i^{(1)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}}\right)$$

where $\mathbf{s}_i^{(l)} = [s_i^{(l)}, s_{i-1}^{(l)}, \cdots, s_{i-L+1}^{(l)}]^T$ and $\hat{\mathbf{H}}_i^{(l)}$ is the RLS estimate of the TV channel matrix associated with the data sequence l.

Suppose that the $\mathbf{s}^{(0)}$ and $\mathbf{s}^{(1)}$ differ for the first time at time instant k_1 (i.e. $s_i^{(0)} = s_i^{(1)},\ i < k_1$ and $s_{k_1}^{(0)} \neq s_{k_1}^{(1)}$). Then we can write the PEP as

$$\begin{aligned}\mathrm{PEP}(\mathbf{s}^{(0)} \rightarrow \mathbf{s}^{(1)}|\boldsymbol{\xi}) &= P\left(\sum_{i=k_1}^{N}\left\|\mathbf{x}_i - \hat{\mathbf{H}}_i^{(0)}\mathbf{s}_i^{(0)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} > \sum_{i=k_1}^{N}\left\|\mathbf{x}_i - \hat{\mathbf{H}}_i^{(1)}\mathbf{s}_i^{(1)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}}\right) \\ &= P\left(\sum_{i=k_1}^{N}\left\|\mathbf{x}_i - \widetilde{\mathbf{B}}_i^{(0)}\hat{\mathbf{h}}_i^{(0)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} > \sum_{i=k_1}^{N}\left\|\mathbf{x}_i - \widetilde{\mathbf{B}}_i^{(1)}\hat{\mathbf{h}}_i^{(1)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}}\right)\end{aligned}$$

where $\hat{\mathbf{h}}_i^{(l)} = \mathrm{vec}(\hat{\mathbf{H}}_i^{(l)})$. Observe that because of the dependence of the RLS estimate $\hat{\mathbf{H}}_i^{(l)}$ on the past values of the sequence $\mathbf{s}^{(l)}$, we still have $\hat{\mathbf{H}}_i^{(0)} \neq \hat{\mathbf{H}}_i^{(1)}$ when the two sequences merge after some time $k_2 \geq k_1 + L$. Thus the PEP depends not only on the length of the error event but also on k_1. This is in contrast with a channel estimate that does not rely on past decisions (such as those based on a training preamble in a time-invariant channel environment).

Let $\boldsymbol{\eta}_i = [\mathbf{h}_i^T, (\hat{\mathbf{h}}_i^{(0)})^T, (\hat{\mathbf{h}}_i^{(1)})^T, \mathbf{z}_i^T]^T$. It can be readily shown [35] that

$$\mathrm{PEP}(\mathbf{s}^{(0)} \rightarrow \mathbf{s}^{(1)}|\boldsymbol{\xi}) = \sum_{\mu_i > 0} \prod_{\substack{j=1 \\ j \neq i}}^{N_o} \frac{\mu_i}{\mu_i - \mu_j}\ . \tag{31}$$

where $N_o = M\,Q\,(3\,L+1)\,(N-k_1+1)$ and $\{\mu_i\}$ are the eigenvalues of $\mathbf{D} = \mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^H]$. Here $\boldsymbol{\eta} = [\boldsymbol{\eta}_{k_1}^T, \boldsymbol{\eta}_{k_1+1}^T, \cdots, \boldsymbol{\eta}_N^T]^T$ and $\mathbf{D}$ is the block diagonal matrix whose iith block matrix is

$$\mathbf{D}_{ii} = \begin{bmatrix} \widetilde{\mathbf{B}}_i^{(0)} & -\widetilde{\mathbf{B}}_i^{(0)} & \mathbf{0} & \mathbf{I} \\ \widetilde{\mathbf{B}}_i^{(0)} & \mathbf{0} & -\widetilde{\mathbf{B}}_i^{(1)} & \mathbf{I} \end{bmatrix}^H \begin{bmatrix} \mathbf{R}_{\mathbf{z}}^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{R}_{\mathbf{z}}^{-1} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{B}}_i^{(0)} & -\widetilde{\mathbf{B}}_i^{(0)} & \mathbf{0} & \mathbf{I} \\ \widetilde{\mathbf{B}}_i^{(0)} & \mathbf{0} & -\widetilde{\mathbf{B}}_i^{(1)} & \mathbf{I} \end{bmatrix}$$

The expression for $\mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^H]$ depends on the true channel model as well as the type of channel estimator used. The PEP is obtained by averaging over the interference statistics $\boldsymbol{\xi}$

$$\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}) = \mathbb{E}[\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}|\boldsymbol{\xi})] \tag{32}$$

Note that this expression is fairly general and is applicable to any adaptive linear estimation scheme. It is clear that the computation of the PEP given in (32) and (31) is prohibitive. To obtain a computationally feasible approximate expression for the PEP, several simplifying assumptions have to be made:

1. The channel estimates are based on correct decision feedback (CDFB).
2. The channel of the desired signal is quasi-static over the length of dominant error events.
3. The true $\mathbf{R}_{\mathbf{z}}$ is known and the CCI is uncorrelated temporally beyond a symbol duration.
4. The channel estimation error is small compared to the colored noise.

The first assumption assures that the channel estimates are based only on the transmitted sequence. This can be justified since it is highly probable that the true transmitted sequence will be among the $|\mathcal{A}|^{L-1}$ survivor paths in the JCD-IS receiver where each state maintains its own survivor and channel estimates. The second assumption is satisfied when considering channel dynamics over the lengths of typical error events. The requirement that $\mathbf{R}_{\mathbf{z}}$ be known can be justified from the fact that under CDFB, the sample covariance $\hat{\mathbf{R}}_{\mathbf{z}}$ of the residual error $\mathbf{x}_i - \mathbf{H}_i\mathbf{s}_i$ forms a good approximation to $\mathbf{R}_{\mathbf{z}}$ when the receiver is in the tracking mode; in fact, $\hat{\mathbf{R}}_{\mathbf{z}} = \mathbf{R}_{\mathbf{z}} + O(1/\sqrt{N})$. The assumption of weak temporal correlation of the colored noise can be justified if the CCI has relatively small delay spread. The final assumption holds approximately under conditions of high SIR and a sufficiently long training sequence.

Under assumption 1, we have $\hat{\mathbf{H}}_i^{(0)} = \hat{\mathbf{H}}_i^{(1)} = \hat{\mathbf{H}}_i$. The approximate PEP conditioned on the true channel, the channel estimates and the CCI data can be written as

$$\begin{aligned}
\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}|\mathbf{H}_i, \hat{\mathbf{H}}_i) &= P\left(\sum_{i=k_1}^{N} \left\|\mathbf{x}_i - \hat{\mathbf{H}}_i\mathbf{s}_i^{(0)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} > \sum_{i=k_1}^{N} \left\|\mathbf{x}_i - \hat{\mathbf{H}}_i\mathbf{s}_i^{(1)}\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}}\right) \\
&= P\left(\sum_{i=k_1}^{k_1+L_e-1} \left(\left\|\hat{\mathbf{H}}_i\boldsymbol{\epsilon}_i\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} + 2\,\mathrm{Re}\{(\mathbf{x}_i - \hat{\mathbf{H}}_i\mathbf{s}_i^{(0)})^H\mathbf{R}_{\mathbf{z}}^{-1}\hat{\mathbf{H}}_i\boldsymbol{\epsilon}_i\}\right) < 0\right) \\
&= P\left(\sum_{i=k_1}^{k_1+L_e-1} \left(\left\|\hat{\mathbf{H}}_i\boldsymbol{\epsilon}_i\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} + 2\,\mathrm{Re}\left\{\left((\mathbf{H}_i - \hat{\mathbf{H}}_i)\mathbf{s}_i^{(0)} + \mathbf{z}_i\right)^H\mathbf{R}_{\mathbf{z}}^{-1}\hat{\mathbf{H}}_i\boldsymbol{\epsilon}_i\right\}\right) < 0\right) \\
&\overset{(a)}{=} P\left(\sum_{i=k_1}^{k_1+L_e-1} \left(\left\|\hat{\mathbf{H}}_i\boldsymbol{\epsilon}_i\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} + 2\,\mathrm{Re}\left\{\mathbf{z}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\hat{\mathbf{H}}_i\boldsymbol{\epsilon}_i\right\}\right) < 0\right) \\
&\overset{(b)}{=} P\left(\sum_{i=k_1}^{k_1+L_e-1} \left(\left\|\mathbf{E}_i\,\hat{\mathbf{h}}_i\right\|^2_{\mathbf{R}_{\mathbf{z}}^{-1}} + 2\,\mathrm{Re}\left\{\mathbf{z}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{E}_i\,\hat{\mathbf{h}}_i\right\}\right) < 0\right)
\end{aligned}$$

where L_e is the length of the error event and $\boldsymbol{\epsilon}_i = \mathbf{s}_i^{(0)} - \mathbf{s}_i^{(1)}$ is the $L \times 1$ error vector. Equality (a) follows from the assumption 4 and equality (b) follows from the identity $\hat{\mathbf{H}}_i \boldsymbol{\epsilon}_i = \mathbf{E}_i \, \hat{\mathbf{h}}_i$ where $\mathbf{E}_i = \boldsymbol{\epsilon}_i^T \otimes \mathbf{I}_{MQ}$ and $\hat{\mathbf{h}}_i = \mathrm{vec}(\hat{\mathbf{H}}_i)$.

Next we find the conditional variance of $2\,\mathrm{Re}\left\{\mathbf{z}_i^H \mathbf{R}_{\mathbf{z}}^{-1} \mathbf{E}_i \, \hat{\mathbf{h}}_i\right\}$.

$$\mathrm{var}\left(2\,\mathrm{Re}\left\{\mathbf{z}_i^H \mathbf{R}_{\mathbf{z}}^{-1} \mathbf{E}_i \, \hat{\mathbf{h}}_i\right\} | \, \hat{\mathbf{h}}_i\right) = 2 \hat{\mathbf{h}}_i^H \mathbf{E}_i^H \mathbf{R}_{\mathbf{z}}^{-1} \mathbf{E}_i \, \hat{\mathbf{h}}_i$$

Here we have assumed that $\hat{\mathbf{h}}_i$ is independent of $\mathbf{z}_i$ as it depends on previous noise samples. By assumption 3 for the CCI[2] , it follows that

$$\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)} | \mathbf{H}_i, \hat{\mathbf{H}}_i) = Q\left(\sqrt{\frac{1}{2} \sum_{i=k_1}^{k_1+L_e-1} \hat{\mathbf{h}}_i^H \mathbf{E}_i^H \mathbf{R}_{\mathbf{z}}^{-1} \mathbf{E}_i \, \hat{\mathbf{h}}_i}\right) \tag{33}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} d\,y$.

The conditional probability density of $\hat{\mathbf{h}}_i$ is

$$f(\hat{\mathbf{h}}_i | \mathbf{h}_i) \sim \mathbb{C}\mathcal{N}\left(\mathbf{h}_i, \mathbf{R}_{\Delta \mathbf{h}_i}\right)$$

where $\mathbf{R}_{\Delta \mathbf{h}_i}$ is the covariance of the channel estimation error vector $\Delta \mathbf{h}_i = \hat{\mathbf{h}}_i - \mathbf{h}_i$. By assumption of a Rayleigh fading channel, the density of $\mathbf{h}_i$ is also Gaussian, that is,

$$f(\mathbf{h}_i) \sim \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{R}_{\mathbf{h}}\right)$$

It is useful to observe that $\mathbf{R}_{\mathbf{h}}$ is completely characterized by the true channel with its associated dynamics while $\mathbf{R}_{\Delta \mathbf{h}_i}$ depends on the type of channel estimator[3]. Under assumption 2 and 4, it can be shown (see Appendix C) that the channel estimation error vector $\Delta \mathbf{h}_i$ is approximately independent of $\mathbf{h}_i$ and independent and identically distributed. It then follows that the probability density of the channel estimator is

$$f(\hat{\mathbf{h}}_i) \sim \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{R}_{\mathbf{h}} + \mathbf{R}_{\Delta \mathbf{h}}\right) \tag{34}$$

Since $\hat{\mathbf{h}}_i = \mathbf{h}_i + \Delta \mathbf{h}_i$, the vector $[\hat{\mathbf{h}}_{k_1}^T, \cdots, \hat{\mathbf{h}}_{k_1+L_e-1}^T]^T$ has covariance $\boldsymbol{\Sigma} = \mathbf{1}_{L_e} \otimes \mathbf{R}_{\mathbf{h}} + \mathbf{I}_{L_e} \otimes \mathbf{R}_{\Delta \mathbf{h}}$ where $\mathbf{1}_p$ is the $p \times p$ matrix of all ones. Let $\boldsymbol{\Phi}_k = \mathbf{E}_k^H \mathbf{R}_{\mathbf{z}}^{-1} \mathbf{E}_k$ and let $\boldsymbol{\Phi}$ be the $L_e\, M\, Q\, L \times L_e\, M\, Q\, L$ block diagonal matrix where

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{k_1} & & & \mathbf{0} \\ & \boldsymbol{\Phi}_{k_1+1} & & \\ & & \ddots & \\ \mathbf{0} & & & \boldsymbol{\Phi}_{k_1+L_e-1} \end{bmatrix}$$

Then it can be shown that the unconditional PEP can be readily calculated as

$$\mathrm{PEP}(\mathbf{s}^{(0)} \to \mathbf{s}^{(1)}) = \sum_{k=1}^{L_e\, M\, Q\, L} \frac{a_k}{2} \left(1 - \sqrt{\frac{\mu_k}{\mu_k + 2}}\right) \tag{35}$$

where $\{\mu_k\}$ are the non-zero distinct eigenvalues of $\frac{1}{2} \boldsymbol{\Sigma}^{H/2} \boldsymbol{\Phi} \boldsymbol{\Sigma}^{1/2}$, $\boldsymbol{\Sigma}^{1/2}$ is the square root matrix of $\boldsymbol{\Sigma}$ such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{H/2}$, and where

$$a_k = \prod_{j \neq k} \frac{1}{1 - \frac{\mu_k}{\mu_j}}$$

[2]Technically speaking, it is unrealistic to assume complete knowledge of the conditional covariance (conditioned on the interference data). Knowledge of the averaged covariance is less restrictive and we use a narrowband CCI assumption (see equation (24)) to ensure the conditional covariance is within a scalar multiple of the averaged covariance.

[3]This includes the type of adaptive algorithm as well as the channel model.

We are interested in the average bit error rate rather than just the PEP. The form of the PEP does not allow us to use the transfer function approach [36] and hence we need to write out the average error rate in terms of the sum of all the pairwise errors. In practice one restricts interest to a small number of error events rather than evaluating the infinite sum. Such a bound called the truncated union bound (TUB) [33] can be written as,

$$P_b \approx \frac{1}{n} \sum_j e_{ij} \mathrm{PEP}(\mathbf{s}^{(i)} \to \mathbf{s}^{(j)}) \tag{36}$$

where n is the number of input bits and e_{ij} is the number of bit errors in the error event.

V Practical Issues

A Complexity of the JCD-IS Receiver

The JCD-IS receiver proposed in section III has a computational complexity that is exponential in the number of states in the trellis, that is $|\mathcal{A}|^{L-1}$ where $|\mathcal{A}|$ denotes the alphabet size. Thus channels with long impulse responses lead to an impractical receiver. The total computational complexity for the update algorithm in Table 1 for each state in the trellis is

$$\begin{aligned}\text{Parameter update complexity} \;=\;& 12(MQ)^3 + 6(M\nu)^3 + 12(MQ)^2(M\nu) + 24(M\nu)^2(MQ) + 8(M\nu)^2 \\ & +18(MQ)^2 + 18(M\nu)(MQ) + 14(MQ) + 2(M\nu) + 4(Q\nu L) + 9 \;\text{ real flops}\end{aligned}$$

where flops indicates the number of floating point operations. The update of the survivor path to each state requires a computational complexity of $|\mathcal{A}|(4Q\nu L + 6(MQ)^2 M\nu + 6(MQ) + 2)$ flops.

Let's illustrate this computational complexity with an example. We assume the following parameters : $\pi/4$-DQPSK modulation ($|\mathcal{A}| = 4$), $M = 2, Q = 2, \nu = 4, L = 4$. The slot length excluding guard and power ramp up time is $N = 156$ symbols. Then the complexity for the receiver per time index is 149.25 MFlops. To further get an idea of the computing power required at a base station to process sequentially each user in a TDMA frame, we take the example of an IS-54/136 air interface. The base station receiver can use the guard and power ramp up time of 6 symbols duration of the next user's time slot to detect the previous user's data. The total slot time is 6.67 ms. Based on this, the required computing power is 22.4 GFlops/s which is beyond the computing power of most major digital signal processing chips. On the other hand, because the algorithm has an inherent parallel processing structure, the computing power can be reduced by using multiple processors simultaneously. For example, if we assume each state in the trellis has its own processor, then the computing power required per DSP chip is 350 MFlops/s. Table 2 shows the total computational complexity of the JCD-IS receiver as a function of both the channel length L and the alphabet size $|\mathcal{A}|$. The number shown here are those of a WRLS scheme not optimized for complexity. Fast algorithms described [25] could be used to reduce complexity. Further reductions in complexity could be obtained by using so-called stochastic gradient algorithms [26] where the algorithm resembles a weighted LMS algorithm. In summary several schemes can be envisaged to make this receiver more practical.

B A Reduced Complexity JCD-IS Receiver

To alleviate the complexity of the receiver, a trellis with a smaller number of states is used instead. Specifically, a reduced trellis could be constructed whose states are formed from the first η ($0 \leq \eta < L$) elements of the CIR. The

	$L=2$	$L=3$	$L=4$	$L=5$	$L=6$
$\|\mathcal{A}\|=2$	0.69	1.39	2.80	5.63	11.33
$\|\mathcal{A}\|=4$	1.38	5.56	22.38	90.08	362.62

Table 2: Computational complexity of the JCD-IS receiver in GFlops/s.

rest of the state information is estimated as part of the unknown parameters associated with that particular state. Such an approach was first advocated in the DDFSE algorithm [23] for known time-invariant channels. Its natural extension to complement the PSP receiver in a TV unknown channel environment is straightforward. The parameter η is user-defined and its extreme values denote specific receiver structures. When $\eta = 0$, the receiver reduces to an adaptive zero-forcing decision feedback equalizer (ZF-DFE) while for $\eta = L - 1$, the receiver becomes an adaptive MLSE receiver and is identical to the receiver proposed in section III. For intermediate values of η, the receiver can be viewed as a receiver with feedback of decisions delayed by $L - 1 - \eta$ for each of the survivor paths in the reduced trellis. It has performance that lies between the full complexity adaptive MLSE and the adaptive ZF-DFE.

In our case, we observe that

$$\begin{aligned} \mathbf{x}_k &= \mathbf{H}_k \mathbf{s}_k + \mathbf{z}_k \\ &= \left[\underbrace{\overline{\mathbf{H}}_k}_{\eta+1} \; \underbrace{\widetilde{\mathbf{H}}_k}_{L-\eta-1} \right] \begin{bmatrix} \overline{\mathbf{s}}_k \\ \widetilde{\mathbf{s}}_k \end{bmatrix} + \mathbf{z}_k \end{aligned}$$

The reduced trellis with $|\mathcal{A}|^\eta$ states is based on $\overline{\mathbf{s}}$ while the partial state (to be estimated) is $\widetilde{\mathbf{s}}$. The new algorithm is similar to that described in Section III with the addition of an extra step which involves the estimation of the partial state information. The partial state for the ith state can be updated as follows:

$$\widetilde{\mathbf{s}}_{k+1}^{(i)} = \begin{bmatrix} \mathbf{e}_{\eta+1}^T \overline{\mathbf{s}}^{Ji} \\ [\mathbf{I}_{L-\eta-2} | \mathbf{0}] \, \widetilde{\mathbf{s}}_k^{(J)} \end{bmatrix}$$

where $\overline{\mathbf{s}}^{Ji}$ denotes the data vector from the Jth state to the ith state in the reduced trellis and $\mathbf{e}_i$ is the unit vector with one in the ith position and zero elsewhere.

The computational complexity of the reduced complexity receiver can be readily determined from that of the full complexity JCD-IS receiver. A factor of $|\mathcal{A}|^{L-1-\eta}$ reduction in computational complexity is attained. As with the full complexity JCD-IS receiver, the reduced complexity JCD-IS receiver is clearly suitable for parallel computers.

C Abrupt Changes in CCI Statistics

One of the most challenging problems facing reliable channel estimation and data detection is when the CCI statistics undergo a sudden change during the tracking mode. This could be due to the appearance of another CCI coming out of a deep fade or the misalignment of the time slots of interfering co-channel mobiles in a neighbouring cell with that of the desired user. While the proposed JCD-IS receiver is not designed specifically to deal with abrupt changes in the CCI, it is inherently more robust to such changes than conventional adaptive MLSE receivers based on tentative past decisions. This is because the use of zero-delay hypothetical decisions based on the best survivors to each state and the adaptive nature of the residual error covariance could potentially mitigate the effects of sudden change in CCI statistics. A detailed analysis of the effect of this phenomenon on the JCD-IS receiver is difficult. However, we provide some simulation results in section VI to show the robustness of the JCD-IS and JCD-MEDD (JCDE with MEDD decoding) receivers.

VI Numerical Results

To investigate the performance of the proposed receivers, we consider the case of a two element antenna array and BPSK data. The time slot length used is 300 symbols, and a symbol period of $T = 40\mu sec$ was used and a training preamble of 20 symbols is assumed. The transmit filter frequency response is assumed to be a raised cosine pulse with 35% roll-off factor where the impulse response is truncated to 4 symbols duration and the carrier frequency is 1 GHz. In the numerical results, ideal time and frequency synchronization were assumed. However, the effect of time and frequency offsets have been studied in [37]. In this it was observed that the structured channel estimator was robust to timing offsets of the order $0.25T$ and frequency offsets of 200Hz. Note that in a time-varying channel the effect of the frequency offset combines with the effects of fading. In particular, for a frequency offset of Δf, the channel impulse response $h(t, \tau)$ has a multiplicative factor of $\exp(j2\pi\Delta ft)$ contributing to the time-variation.

The TV channel used in the simulations is based on a *discrete* multipath channel model, that is the propagation channel for the ith antenna is

$$\mathbf{c}_i(t, \tau) = \sum_{p=1}^{P} \alpha_p(t) a_i(\theta_p)\delta(\tau - \tau_p) \tag{37}$$

where P is the number of paths, $\alpha_i(t)$ is the complex fading associated with the ith path, $a_i(\theta_p)$ is the complex antenna response of the ith antenna to a signal from direction θ_p.

For the simulation, we use a two ray model with a delay spread of one symbol period for both the desired user and the CCI. The angles of the user are $\theta_1 = -20^o, \theta_2 = 25^o$ while the angles of the CCI are $\theta_1^i = 50^o, \theta_2^i = 70^o$. The fading coefficients for both user and CCI are uncorrelated complex zero mean Gaussian random variables each with a covariance given by $p_l\, J_0(\omega_c v|t - s|/c)$. Here ω_c is the carrier frequency in rad/s. We set the average path strengths to be equal ($p_1 = p_2$) for both user and interference and fix the average channel power of each path of the multipath to be unity. Taking into account the transmit filter and the channel delay spread, the total channel length is $L = 5$. We also set $\eta = 4$ and use an oversampling factor $Q = 2$. The mobile speed is 100 kmh which corresponds to $f_D T = 3.6 \times 10^{-3}$, the forgetting factor λ is set to 0.95 and $w = 0.99$. We shall investigate the receiver performance against CCI at a SNR of 20 dB. These results are averaged over 4800 realizations of the fading and noise processes (60 fade runs $\times$ 80 noise runs). The signal to interference ratio (SIR) plotted in the figures refer to *average* values. As the channel is time-varying (both for the desired signal and the interferer) this could very well mean that in a particular packet (or even part of the packet) the interferer could much stronger than the desired signal, depending on the fading realizations. The tracking performance of the structured channel estimator compared to the unstructured channel estimator has been studied for white noise environments in [22]. It has been demonstrated that mean-squared error gain of about 4dB can be achieved in such environments [22].

Figure 1 shows that the error rate performance of the JCD-IS receivers can be improved if a colored Gaussian (Mahalanobis distance) metric with a structured channel estimator is used instead of a minimum euclidean distance decoding (MEDD) metric and unstructured channel estimator. At a target BER of 1%, using a Mahalanobis distance metric with a structured channel estimator leads to a SIR improvement of about 8 dB over the the unstructured TV channel estimator. This figure demonstrates the advantage of incorporating transmit pulse shape knowledge into the channel description. In the CSI case examined in these figures, only the channel state of the desired user is known and the noise covariance matrix of the interferer (which is time-varying) is estimated. Hence it serves as the lower bound of performance for any estimator. For a given estimator, the lower bound of performance could be obtained when we adapt the estimators with the correct decisions. Figure 2 shows the BER of the JCD-IS receivers using correct decision feedback (CDFB). We observe that the JCD-IS receiver using a structured channel model can achieve error rates that are slightly better that the CDFB receiver using an unstructured channel model. The performance of the JCD-IS algorithm is similar to the adaptive interference suppression algorithm in [17], but the results presented are at a slightly higher $f_D T$, (*i.e.,* time-variation) and a frequency selective interferer.

Figure 3 shows the analytically computed probability of error using the the approximate PEP expression in section IV. This is done by truncating the length of error events to about 50 symbols and error sequences with Hamming weight not more than 7. This is compared with both the JCD-IS receiver and the CDFB based receiver using structured channel estimators. Note that in section IV, the approximation was based on CDFB and narrowband interference assumptions. The expressions were therefore an approximation to the probability of error and *not* an upper bound to the JCD-IS performance. However, as it is based on CDFB assumption, one would expect it to be between the CDFB curve and the JCD-IS curve as observed in Figure 3.

Next, we investigate the BER performance of the proposed reduced complexity DDFSE JCD-IS receiver. From Figure 4, we observe that the BER performance for both the reduced complexity receiver and the full complexity receiver are quite close. For high SIR values, the full complexity receiver has a slightly lower BER than the reduced complexity receiver as expected.

We end this section with a simulation scenario where the JCD-IS receiver operates in an environment when the CCI and the user are slot-misaligned by half a time slot. This corresponds to a change in the CCI statistics in the middle of the slot where the receiver is in the tracking mode. The first CCI has the original parameters as in the first example. The second independent CCI which appears in the middle of the slot at which the first CCI disappears has the following parameters : $\theta_1^i = 45°, \theta_2^i = 65°$. Figure 5 shows the BER of the JCD-IS (structured) and JCD-IS (unstructured) receivers. Both receivers have similar performances for low SIR but the JCD-IS (structured) receiver marginally better than the JCD-IS (unstructured) receiver at high SIR. The performance of the receiver is quite good, if we take into account the fact that we have an abrupt change in the interference statistics. However, as expected, both receivers do not perform as well as when there is only one and the same CCI throughout the slot.

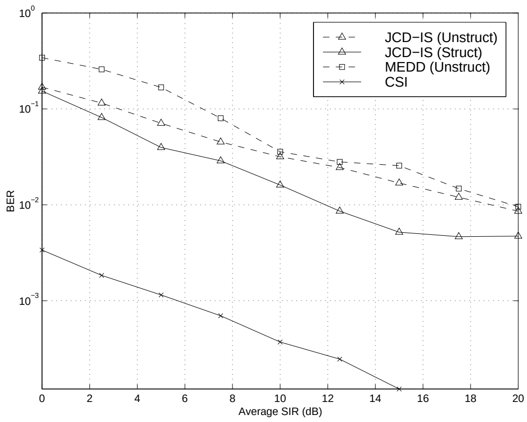

Figure 1: Comparison of BER performances between JCD-IS and MEDD JCD receivers.

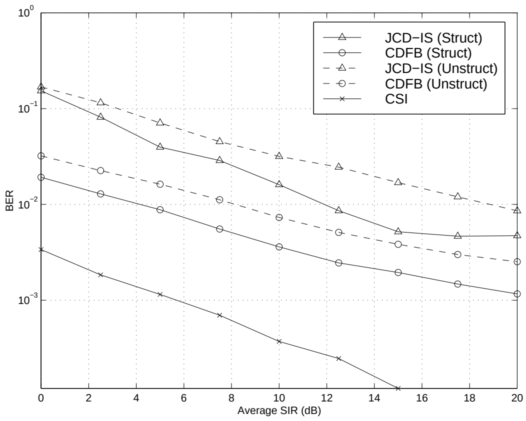

Figure 2: BER performances of JCD-IS receiver and CDFB-based JCD-IS receiver.

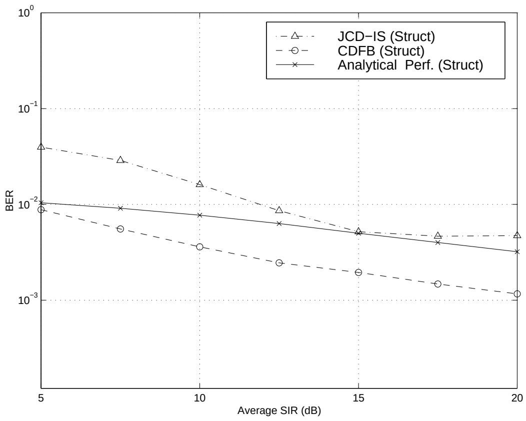

Figure 3: Comparison of analytically computed BER with the JCD-IS receiver and the CDFB-based JCD-IS receiver.

VII Concluding Remarks

By identifying the noise covariance matrix and incorporating it into the decoding metric we effectively perform detection after suppressing the interference subspace. The results obtained in this paper show that this is a promising approach. There are several problems that still need to be addressed. Firstly, new cost criteria for the identification algorithm can be developed. Local minima would be eliminated if a convex cost criterion is chosen. Secondly, a better analysis for the mismatched PEP could be developed without the assumption 4 in Section IV. Next, reduced complexity schemes need to be investigated. Better schemes to detect and handle abrupt changes in the CCI statistics (due to asynchronous interference) need to be developed. Finally, extensions of this idea to CDMA could be investigated.

ACKNOWLEDGEMENT

We gratefully acknowledge stimulating discussions with Bijit Halder. Insightful comments from Prof. Donald Cox greatly improved the presentation of this paper. We also thank David Gesbert, Alle-Jan Van Der Veen and Robert Heath for their comments on the paper. Finally, we thank the anonymous reviewers, whose detailed comments vastly

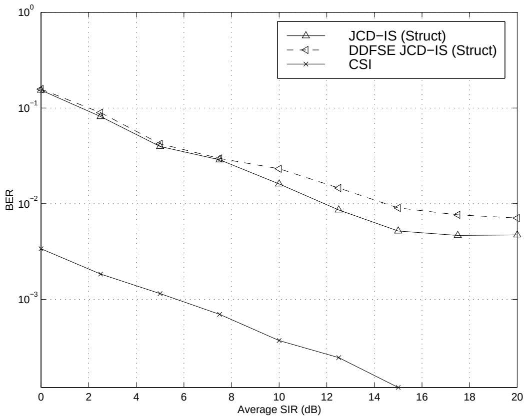

Figure 4: BER performance of the reduced complexity DDFSE JCD-IS receiver.

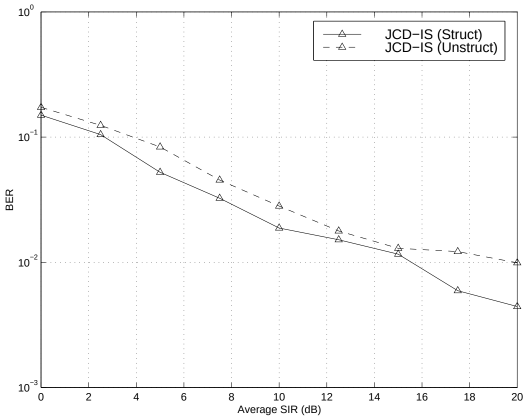

Figure 5: BER performance of the JCD-IS and JCD-MEDD receivers with abrupt change in CCI in the middle of user time slot.

improved the presentation of the paper.

Appendix A Calculation of $\mathbf{V}_{22}$

We define

$$\frac{\partial \mathrm{vec}(\mathbf{U})}{\partial \mathrm{vec}(\mathbf{R})^T} = [\mathrm{vec}(\frac{\partial(\mathbf{U})}{\partial \mathbf{R}_{11}}), \ldots, \mathrm{vec}(\frac{\partial(\mathbf{U})}{\partial \mathbf{R}_{1N}}), \ldots, \mathrm{vec}(\frac{\partial(\mathbf{U})}{\partial \mathbf{R}_{N1}}), \ldots, \mathrm{vec}(\frac{\partial(\mathbf{U})}{\partial \mathbf{R}_{NN}})] \tag{A.1}$$

where $\mathbf{R}$ is a $N \times N$ matrix, and $\mathbf{R}_{mn}$ denotes its $(m,n)^{th}$ element. To calculate $\mathbf{V}_{22}$ we need the following lemma.

Lemma A.1

$$\frac{\partial \mathrm{vec}(\mathbf{R}^{-1})}{\mathrm{vec}(\mathbf{R})^T} = -\mathbf{R}^{-1} \otimes \mathbf{R}^{-1} \tag{A.2}$$

Proof: As $\mathbf{R}^{-1}\mathbf{R} = \mathbf{I}$, we have by differentiating w.r.t. $\mathbf{R}_{mn}$,

$$\frac{\partial \mathbf{R}^{-1}}{\partial \mathbf{R}_{mn}} = -\mathbf{R}^{-1}\mathbf{E}_{mn}\mathbf{R}^{-1} = -(\mathbf{R}^{-1}\mathbf{e}_m)(\mathbf{R}^{-1}\mathbf{e}_n)^T \tag{A.3}$$

where $\mathbf{e}_m$ is the m^{th} unit vector. Hence as,

$$\text{vec}(\frac{\partial \mathbf{R}^{-1}}{\partial \mathbf{R}_{mn}}) = -(\mathbf{R}^{-1}\mathbf{e}_n) \otimes (\mathbf{R}^{-1}\mathbf{e}_m) \tag{A.4}$$

and using (A.1) we get the desired result. ■

Claim A.1

$$\mathbf{V}_{22}|_{\hat{\mathbb{E}}_k[\boldsymbol{\alpha}\boldsymbol{\alpha}^T]=\mathbf{R}} = \mathbf{R}^{-1} \otimes \mathbf{R}^{-1} \tag{A.5}$$

Proof: From (13) we get,

$$\mathbf{V}_{22} = -\frac{\partial(\text{vec}(\mathbf{R}^{-1}\hat{\mathbb{E}}_k[\boldsymbol{\alpha}\boldsymbol{\alpha}^T]\mathbf{R}^{-1}))}{\partial \text{vec}(\mathbf{R})^T} + \frac{\partial(\text{vec}(\mathbf{R}^{-1}))}{\partial \text{vec}(\mathbf{R})^T} \tag{A.6}$$

Clearly the second term of (A.6) is given by Lemma A.1. For the first term we have,

$$\begin{aligned} \frac{\partial(\mathbf{R}^{-1}\hat{\mathbb{E}}_k[\boldsymbol{\alpha}\boldsymbol{\alpha}^T]\mathbf{R}^{-1})}{\partial \mathbf{R}_{mn}} &= \hat{\mathbb{E}}_k[\frac{\partial(\mathbf{R}^{-1})}{\partial \mathbf{R}_{mn}}\boldsymbol{\alpha}\boldsymbol{\alpha}^T\mathbf{R}^{-1} + \mathbf{R}^{-1}\boldsymbol{\alpha}\boldsymbol{\alpha}^T\frac{\partial(\mathbf{R}^{-1})}{\partial \mathbf{R}_{mn}}] \\ &\overset{(a)}{=} -\hat{\mathbb{E}}_k[(\mathbf{R}^{-1}\mathbf{e}_m)(\mathbf{R}^{-1}\mathbf{e}_n)^T\boldsymbol{\alpha}\boldsymbol{\alpha}^T\mathbf{R}^{-1} + \mathbf{R}^{-1}\boldsymbol{\alpha}\boldsymbol{\alpha}^T(\mathbf{R}^{-1}\mathbf{e}_m)(\mathbf{R}^{-1}\mathbf{e}_n)^T] \end{aligned} \tag{A.7}$$

where (a) follows from (A.3). When we evaluate (A.7) for $\hat{\mathbb{E}}_k[\boldsymbol{\alpha}\boldsymbol{\alpha}^T] = \mathbf{R}$ we obtain,

$$\text{vec}(\frac{\partial(\mathbf{R}^{-1}\hat{\mathbb{E}}_k[\boldsymbol{\alpha}\boldsymbol{\alpha}^T]\mathbf{R}^{-1})}{\partial \mathbf{R}_{mn}}) = -2(\mathbf{R}^{-1}\mathbf{e}_n) \otimes (\mathbf{R}^{-1}\mathbf{e}_m) \tag{A.8}$$

Using (A.8), Lemma A.1 and (A.1) we get the desired result. ■

Appendix B Covariance matrix of parametric vector channel

In this appendix, we shall derive the covariance matrix of the vector channel based on the parametric channel model where the propagation channel is composed of discrete multipaths arriving at the receiver array with different angles of arrival and delays. We start by writing the multichannel receiver output as

$$\mathbf{x}(t) = \sum_n s_n \left(\sum_{p=1}^{P} \mathbf{a}(\theta_p)\alpha_p(t)g(t - nT - \tau_p) \right) \tag{B.1}$$

By fractionally sampling (B.1) to obtain the sufficient statistics at a rate Q/T, we obtain

$$\mathbf{x}(kT + qT/Q) = \sum_{m=0}^{L-1} s_{k-m}\mathbf{h}(kT + qT/Q, mT)$$

where

$$\begin{aligned} \mathbf{h}(kT + qT/Q, mT) &= \sum_{p=1}^{P} \mathbf{a}(\theta_p)\alpha_p(kT + qT/Q)g(mT + qT/Q - \tau_p) \\ &= \mathbf{A}(\boldsymbol{\theta})\mathbf{D}(kT + qT/Q)\mathbf{g}(mT + qT/Q, \boldsymbol{\tau}) & \text{(B.2)} \\ &= \left[\mathbf{g}^T(mT + qT/Q, \boldsymbol{\tau}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\alpha}(kT + qT/Q) & \text{(B.3)} \end{aligned}$$

The definitions for the terms in (B.2)-(B.3) are as follows

$$\begin{aligned}
\mathbf{A}(\boldsymbol{\theta}) &= [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_P)] \\
\mathbf{D}(kT + qT/Q) &= \mathrm{diag}[\alpha_1(kT+qT/Q), \alpha_2(kT+qT/Q), \cdots, \alpha_P(kT+qT/Q)] \\
\mathbf{g}(mT + qT/Q, \boldsymbol{\tau}) &= \begin{bmatrix} g(mT+qT/Q-\tau_1) \\ g(mT+qT/Q-\tau_2) \\ \vdots \\ g(mT+qT/Q-\tau_P) \end{bmatrix} \\
\boldsymbol{\alpha}(kT+qT/Q) &= [\alpha_1(kT+qT/Q), \alpha_2(kT+qT/Q), \cdots, \alpha_P(kT+qT/Q)]^T
\end{aligned}$$

and where $\diamond$ is the Khatri-Rao (column-wise) matrix product (see Appendix D).

Our goal is to find the covariance matrix of $\mathbf{h}_k$. It can be verified that $\mathbf{h}_k = \mathrm{vec}(\mathcal{H}_k^T)$ where

$$\mathcal{H}_k = \begin{bmatrix} \left[\mathbf{g}^T(0,\boldsymbol{\tau}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\alpha}(kT) & \cdots & \left[\mathbf{g}^T((Q-1)T/Q,\boldsymbol{\tau}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\alpha}(kT+(Q-1)T/Q) \\ \left[\mathbf{g}^T(T,\boldsymbol{\tau}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\alpha}(kT) & \cdots & \left[\mathbf{g}^T(T+(Q-1)T/Q,\boldsymbol{\tau}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\alpha}(kT+(Q-1)T/Q) \\ \vdots & \vdots & \vdots \\ \left[\mathbf{g}^T((L-1)T,\boldsymbol{\tau}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\alpha}(kT) & \cdots & \left[\mathbf{g}^T((L-1)T+(Q-1)T/Q,\boldsymbol{\tau}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \boldsymbol{\alpha}(kT+(Q-1)T/Q) \end{bmatrix}$$

Let

$$\mathbf{G}_i = \begin{bmatrix} \mathbf{g}^T(iT/Q,\boldsymbol{\tau}) \\ \mathbf{g}^T(T+iT/Q,\boldsymbol{\tau}) \\ \cdots \\ \mathbf{g}^T((L-1)T+iT/Q,\boldsymbol{\tau}) \end{bmatrix}$$

This definition allows us to write

$$\mathcal{H}_k = [\,[\mathbf{G}_0 \diamond \mathbf{A}(\boldsymbol{\theta})]\boldsymbol{\alpha}(kT),\ [\mathbf{G}_1 \diamond \mathbf{A}(\boldsymbol{\theta})]\boldsymbol{\alpha}(kT+T/Q),\ \cdots, [\mathbf{G}_{Q-1} \diamond \mathbf{A}(\boldsymbol{\theta})]\boldsymbol{\alpha}(kT+(Q-1)T/Q)\,]$$

Recall $\mathbf{h}_k = \mathrm{vec}(\mathcal{H}_k^T) = \mathbf{P}_1 \mathrm{vec}(\mathcal{H}_k)$ where $\mathbf{P}_1$ is a permutation matrix that reorders the rows of $\mathrm{vec}(\mathcal{H}_k)$ and is given by

$$\mathbf{P}_1 = \begin{bmatrix} \mathbf{I}_Q \otimes \mathbf{e}_1^T \\ \vdots \\ \mathbf{I}_Q \otimes \mathbf{e}_{ML}^T \end{bmatrix}_{MQL \times MQL}$$

where $\mathbf{e}_i$ is the $ML \times 1$ unit vector with a one in the ith position and zero elsewhere.

Now

$$\mathrm{vec}(\mathcal{H}_k) = \mathbf{F} \begin{bmatrix} \boldsymbol{\alpha}(kT) \\ \boldsymbol{\alpha}(kT+T/Q) \\ \vdots \\ \boldsymbol{\alpha}(kT+(Q-1)T/Q) \end{bmatrix} \tag{B.4}$$

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{G}_0 \diamond \mathbf{A}(\boldsymbol{\theta}) & & & \\ & \mathbf{G}_1 \diamond \mathbf{A}(\boldsymbol{\theta}) & & \\ & & \ddots & \\ & & & \mathbf{G}_{Q-1} \diamond \mathbf{A}(\boldsymbol{\theta}) \end{bmatrix}_{MLQ \times PQ}$$

By defining another $PQ \times PQ$ permutation matrix $\mathbf{P}_2$

$$\mathbf{P}_2 = \begin{bmatrix} \mathbf{I}_P \otimes \mathbf{e}_1^T \\ \vdots \\ \mathbf{I}_P \otimes \mathbf{e}_Q^T \end{bmatrix}_{PQ \times PQ}$$

where $\mathbf{e}_i$ now denotes a $Q \times 1$ unit vector such that

$$\mathbf{P}_2 \begin{bmatrix} \boldsymbol{\alpha}(kT) \\ \boldsymbol{\alpha}(kT + T/Q) \\ \vdots \\ \boldsymbol{\alpha}(kT + (Q-1)T/Q) \end{bmatrix} = \boldsymbol{\alpha}_k \tag{B.5}$$

In (B.5), $\boldsymbol{\alpha}_k$ denotes the fading variables arranged in order of the path index first, i.e.

$$\boldsymbol{\alpha}_k = [\alpha_1(kT), \alpha_1(kT + T/Q), \cdots, \alpha_1(kT + (Q-1)T/Q), \cdots, \alpha_P(kT), \alpha_P(kT + T/Q), \cdots, \alpha_P(kT + (Q-1)T/Q)]^T$$

Using (B.4) and (B.5), we obtain the linear model for $\mathbf{h}_k$ as a function of $\boldsymbol{\alpha}_k$

$$\mathbf{h}_k = \mathbf{P}_1 \mathbf{F} \mathbf{P}_2 \boldsymbol{\alpha}_k \tag{B.6}$$

The cross-covariance matrix of $\mathbf{h}_k$ can now be calculated as

$$\mathbb{E}(\mathbf{h}_k \mathbf{h}_{k+n}^H) = \mathbf{P}_1 \mathbf{F} \mathbf{P}_2 \mathbf{R}_{\boldsymbol{\alpha}_k \boldsymbol{\alpha}_{k+n}} \mathbf{P}_2^T \mathbf{F}^H \mathbf{P}_1^T \tag{B.7}$$

For the case of independent Rayleigh fading between paths,

$$\mathbf{R}_{\boldsymbol{\alpha}_k \boldsymbol{\alpha}_{k+n}} = \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_n \tag{B.8}$$

where $\boldsymbol{\Lambda}$ is the diagonal $P \times P$ matrix containing the powers of the multipaths and the $Q \times Q$ matrix $\boldsymbol{\Gamma}_n$ describes the temporal correlation of each path at lag nT and its ijth component is

$$[\boldsymbol{\Gamma}_n]_{ij} = J_0(\omega_D(|n| + |i-j|/Q))$$

where $\omega_D = 2\pi f_d T$ is the normalized Doppler angular frequency and f_d is the Doppler frequency in Hz. Using (B.7) and (B.8), the cross-covariance matrix is given by

$$\mathbf{R}_{\mathbf{h}_k \mathbf{h}_{k+n}} = \mathbf{P}_1 \mathbf{F} \mathbf{P}_2 (\boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_n) \mathbf{P}_2^T \mathbf{F}^H \mathbf{P}_1^T \tag{B.9}$$

This completes the derivation of the (cross)-covariance matrix of the $\mathbf{h}_k$.

Appendix C Covariance matrix of channel estimation error

In this appendix, we shall investigate the channel estimation error vector for the RLS channel updates. Our objective is to obtain an (asymptotic) expression for the covariance matrix of the channel estimation error vector for the purpose of computing the PEP. We shall also show that the channel estimation error vector is approximately independent of the true channel vector under quasi-static channel conditions over the length of typical error events.

The channel estimation error vector is defined as

$$\Delta\mathbf{h}_k = \hat{\mathbf{h}}_k - \mathbf{h}_k = \underbrace{\left(\hat{\mathbf{h}}_k - \mathbb{E}\hat{\mathbf{h}}_k\right)}_{\overline{\Delta\mathbf{h}}_k} + \underbrace{\left(\mathbb{E}\hat{\mathbf{h}}_k - \mathbf{h}_k\right)}_{\widetilde{\Delta\mathbf{h}}_k}$$

The vector $\overline{\Delta\mathbf{h}}_k$ is the channel estimation noise vector and it quantifies the deviation of the channel estimate from its mean value as a result of noise perturbations. On the other hand, the vector $\widetilde{\Delta\mathbf{h}}_k$ is the channel lag vector and is used to quantify the tracking error inherent in the adaptive algorithm as a result of channel dynamics. We shall investigate these quantities separately. We shall also see that both vectors are statistically independent so that the sum of their covariance matrices gives the total covariance matrix of $\Delta\mathbf{h}_k$.

C.1 Channel estimation noise vector $\overline{\Delta\mathbf{h}}_k$

The RLS channel estimate is

$$\hat{\mathbf{h}}_k = \overline{\mathbf{G}}\hat{\mathbf{c}}_k = \overline{\mathbf{G}}\left[\sum_i \lambda^{k-i}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\right]^{-1}\left[\sum_i \lambda^{k-i}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{x}_i\right] \tag{C.1}$$

Let $\mathbf{\Omega}_k = \left[\sum_i \lambda^{k-i}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\right]^{-1}$. Since $\mathbf{x}_i = \mathbf{B}_i\mathbf{c}_i + \mathbf{z}_i$ and taking expectation of (C.1) over the noise,

$$\mathbb{E}\hat{\mathbf{h}}_k = \overline{\mathbf{G}}\mathbf{\Omega}_k\sum_i \lambda^{k-i}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\mathbf{c}_i$$

The channel estimation noise vector can now be written as

$$\overline{\Delta\mathbf{h}}_k = \overline{\mathbf{G}}\mathbf{\Omega}_k\sum_i \lambda^{k-i}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{z}_i \tag{C.2}$$

It is clear from (C.2) that $\overline{\Delta\mathbf{h}}_k$ has zero mean. The covariance matrix is

$$\mathrm{Cov}\left(\overline{\Delta\mathbf{h}}_k\right) = \mathbb{E}\left(\overline{\Delta\mathbf{h}}_k\ \overline{\Delta\mathbf{h}}_k^H\right) = \overline{\mathbf{G}}\mathbf{\Omega}_k\left(\sum_i \lambda^{2(k-i)}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\right)\mathbf{\Omega}_k\overline{\mathbf{G}}^T \tag{C.3}$$

where we have made use of weak temporal correlation of $\mathbf{z}_i$ in assumption 3 in section IV.B to obtain the second equality. Since we are concerned with the stationary properties of the channel error statistics in the tracking mode and assuming ergodicity, the large sample limit (asymptotic) covariance matrix is sought. Hence

$$\begin{aligned}
\mathbf{R}_{\overline{\Delta\mathbf{h}}} &= \lim_{k\to\infty}\overline{\mathbf{G}}\mathbf{\Omega}_k\left(\sum_i \lambda^{2(k-i)}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\right)\mathbf{\Omega}_k\overline{\mathbf{G}}^T \\
&= \overline{\mathbf{G}}\mathbf{\Omega}\left(\lim_{k\to\infty}\sum_i \lambda^{2(k-i)}\mathbf{B}_i^H\mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\right)\mathbf{\Omega}\overline{\mathbf{G}}^T
\end{aligned}$$

where $\mathbf{\Omega} = \lim_{k\to\infty} \mathbf{\Omega}_k$. By the continuity of the matrix inverse mapping and provided the limit exist,

$$
\begin{aligned}
\mathbf{\Omega} &= \lim_{k\to\infty}\left[\sum_i^k \lambda^{k-i}\mathbf{B}_i^H \mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\right]^{-1} \\
&= \left[\lim_{k\to\infty}\sum_i^k \lambda^{k-i}\mathbf{B}_i^H \mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i\right]^{-1} \\
&= \left[\lim_{k\to\infty}\sum_i^k \lambda^{k-i}\overline{\mathbf{G}}^T\left(\mathbf{s}_i^* \otimes \mathbf{I}_{MQ}\right)\mathbf{R}_{\mathbf{z}}^{-1}\left(\mathbf{s}_i^T \otimes \mathbf{I}_{MQ}\right)\overline{\mathbf{G}}\right]^{-1} \\
&\overset{(a)}{=} \left[\lim_{k\to\infty}\sum_i^k \lambda^{k-i}\overline{\mathbf{G}}^T\left(\mathbf{s}_i^*\mathbf{s}_i^T \otimes \mathbf{R}_{\mathbf{z}}^{-1}\right)\overline{\mathbf{G}}\right]^{-1} \\
&\overset{(b)}{=} \left[\frac{1}{1-\lambda}\overline{\mathbf{G}}^T\left(\mathbb{E}(\mathbf{s}^*\mathbf{s}^T) \otimes \mathbf{R}_{\mathbf{z}}^{-1}\right)\overline{\mathbf{G}}\right]^{-1} \\
&\overset{(c)}{=} \left[\frac{E_s}{1-\lambda}\overline{\mathbf{G}}^T\left(\mathbf{I}_L \otimes \mathbf{R}_{\mathbf{z}}^{-1}\right)\overline{\mathbf{G}}\right]^{-1} \qquad \text{(C.4)}
\end{aligned}
$$

where equality (a) follows from the identity (D.3) in Appendix D, equality (b) follows from ergodicity[4] and (c) follows from an IID assumption of the user data with $\mathbb{E}|s_k|^2 = E_s$. It is readily shown that

$$
\lim_{k\to\infty}\sum_i^k \lambda^{2(k-i)}\mathbf{B}_i^H \mathbf{R}_{\mathbf{z}}^{-1}\mathbf{B}_i = \frac{E_s}{1-\lambda^2}\overline{\mathbf{G}}^T\left(\mathbf{I}_L \otimes \mathbf{R}_{\mathbf{z}}^{-1}\right)\overline{\mathbf{G}} \qquad \text{(C.5)}
$$

Using (C.4) and (C.5), we have

$$
\mathbf{R}_{\overline{\Delta\mathbf{h}}} = \frac{1-\lambda}{E_s(1+\lambda)}\overline{\mathbf{G}}\left[\overline{\mathbf{G}}^T\left(\mathbf{I}_L \otimes \mathbf{R}_{\mathbf{z}}^{-1}\right)\overline{\mathbf{G}}\right]^{-1}\overline{\mathbf{G}}^T \qquad \text{(C.6)}
$$

C.2 Channel lag error vector $\widetilde{\Delta\mathbf{h}}_k$

In this section, we shall derive the asymptotic covariance matrix of the channel lag error vector $\widetilde{\Delta\mathbf{h}}_k$. The approach taken here follows closely the tracking error analysis in [38]. If the variations of the channel is slow compared to the memory of the algorithm, then it can be shown that [38]

$$
\widetilde{\Delta\mathbf{h}}_k = \lambda\,\widetilde{\Delta\mathbf{h}}_{k-1} - (\mathbf{h}_k - \mathbf{h}_{k-1}) \qquad \text{(C.7)}
$$

Taking the z-transform of (C.7) gives

$$
\widetilde{\Delta\mathbf{h}}(z) = \frac{z^{-1}-1}{1-\lambda z^{-1}}\mathbf{h}(z) \qquad \text{(C.8)}
$$

We can interpret equations (C.7) and (C.8) as the channel lag error vector being the output of a linear time-invariant filter. The power spectrum of $\widetilde{\Delta\mathbf{h}}$ can now be calculated as

$$
\mathbf{S}_{\widetilde{\Delta\mathbf{h}}}(z) = \left(\frac{z^{-1}-1}{1-\lambda z^{-1}}\right)\left(\frac{z-1}{1-\lambda z}\right)\mathbf{S}_{\mathbf{h}}(z) \qquad \text{(C.9)}
$$

where $\mathbf{S}_{\mathbf{h}}(z)$ is the power spectrum of the true channel.

[4] A factor of $\frac{1}{k}$ has been implicitly assumed to be included in (a) before taking the limit. This factor cancels out in the expression for $\mathrm{Cov}\left(\overline{\Delta\mathbf{h}}_k\right)$ and does not affect the final asymptotic expression.

Using (C.9), the cross-covariance matrix of $\widetilde{\Delta \mathbf{h}}$ can be readily calculated as

$$\mathbb{E}\left(\widetilde{\Delta \mathbf{h}}_k \widetilde{\Delta \mathbf{h}}_{k+n}^H\right) = \frac{1}{2\,\pi} \int_{-\pi}^{\pi} \frac{2 - 2\,\cos\omega}{1 + \lambda^2 - 2\,\lambda\,\cos\omega} \mathbf{S}_{\mathbf{h}}(e^{j\omega})\, e^{j\,n\,\omega}\; d\omega \tag{C.10}$$

In particular, we shall find the covariance (at zero lag) of $\widetilde{\Delta \mathbf{h}}_k$ under a parametric multipath Rayleigh fading channel model. Using (B.9), it is easily established that

$$\mathbf{S}_{\mathbf{h}}(e^{j\omega}) = \mathbf{P}_1 \mathbf{F} \mathbf{P}_2 \left(\mathbf{\Lambda} \otimes \mathbf{\Gamma}(e^{j\omega})\right) \mathbf{P}_2^T \mathbf{F}^H \mathbf{P}_1^T$$

where

$$\mathbf{\Gamma}_{mn}(e^{j\omega}) = \begin{cases} \frac{2\,e^{j\omega(|m-n|/Q)}}{\omega_D \sqrt{1-\omega^2/\omega_D^2}} & |\omega| < \omega_D \\ 0 & |\omega| \geq \omega_D \end{cases} \qquad m, n = 1, 2, \cdots, Q. \tag{C.11}$$

and where $\omega_D = 2\,\pi\,f_d\,T$ is the normalized Doppler angular frequency. Using (C.11), the covariance matrix of $\widetilde{\Delta \mathbf{h}}_k$ can now be found as

$$\mathbf{R}_{\widetilde{\Delta \mathbf{h}}} = \mathbf{P}_1 \mathbf{F} \mathbf{P}_2 \left(\mathbf{\Lambda} \otimes \left(\frac{1}{2\,\pi} \int_{-\pi}^{\pi} \frac{2 - 2\,\cos\omega}{1 + \lambda^2 - 2\,\lambda\,\cos\omega} \mathbf{\Gamma}(e^{j\omega})\; d\omega\right)\right) \mathbf{P}_2^T \mathbf{F}^H \mathbf{P}_1^T \tag{C.12}$$

C.3 Total channel estimation error covariance

Using (C.2) and (C.7), it is easy to see that $\overline{\Delta \mathbf{h}}_k$ and $\widetilde{\Delta \mathbf{h}}_k$ are independent. The total channel estimation error covariance is thus given by the sum of their covariances

$$\mathbf{R}_{\Delta \mathbf{h}} = \mathbf{R}_{\overline{\Delta \mathbf{h}}} + \mathbf{R}_{\widetilde{\Delta \mathbf{h}}} \tag{C.13}$$

where $\mathbf{R}_{\overline{\Delta \mathbf{h}}}$ and $\mathbf{R}_{\widetilde{\Delta \mathbf{h}}}$ are given by (C.6) and (C.12) respectively. This completes the derivation of the channel estimation error covariance matrix. Equation (C.7) indicates that the total channel estimation error vector $\Delta \mathbf{h}_k$ is not temporally independent. However assumption 4 in Section IVB implies that $\overline{\Delta \mathbf{h}}$ dominates $\widetilde{\Delta \mathbf{h}}$ and thus we can approximate $\Delta \mathbf{h}_k$ as an IID process.

Appendix D Results on Kronecker products

The results summarized here can be found in [39], Table II (T2.4,T2.6,T2.13,T2.14). Here $\mathbf{R}$ is a real symmetric matrix.

$$(\mathbf{R}^{-1} \otimes \mathbf{R}^{-1})^{-1} = \mathbf{R} \otimes \mathbf{R} \tag{D.1}$$

$$(\mathbf{R} \otimes \mathbf{R})\mathrm{vec}(\mathbf{U}) = \mathrm{vec}(\mathbf{RUR}) \tag{D.2}$$

$$(\mathbf{A} \otimes \mathbf{D})(\mathbf{F} \otimes \mathbf{U}) = (\mathbf{AF}) \otimes (\mathbf{DU}) \tag{D.3}$$

If the eigenvalues of $\mathbf{M}$ are $\{\psi_k\}$ and the eigenvalues of $\mathbf{N}$ are $\{\nu_k\}$ then we have,

$$\mathrm{eig}(\mathbf{N} \otimes \mathbf{M}) = \psi_i \nu_k \tag{D.4}$$

If $\mathbf{U}$ is $(t \times u)$ and $\mathbf{F}$ is $(q \times u)$, the Khatri-Rao product is denoted by $\mathbf{F} \diamond \mathbf{U}$ and is defined [39] as,

$$\mathbf{F} \diamond \mathbf{U} \stackrel{\triangle}{=} [\mathbf{F}_1 \otimes \mathbf{U}_1, \mathbf{F}_2 \otimes \mathbf{U}_2, \ldots, \mathbf{F}_u \otimes \mathbf{U}_u] \tag{D.5}$$

where $\mathbf{F}_i$ denotes the i^{th} column of matrix $\mathbf{F}$. The property used from [39] (Table III, T3.13),

$$\mathrm{vec}(\mathbf{AVD}) = (\mathbf{D}^T \diamond \mathbf{A})\mathrm{vecd}(\mathbf{V}) \tag{D.6}$$

if $\mathbf{V}$ is diagonal and vecd$(\cdot)$ denotes the vector formed by the diagonal elements of the matrix.

References

[1] G. D. Forney, "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Transactions on Information Theory*, vol. 18, pp. 363–378, May 1972.

[2] S. U. H. Qureshi, "Adaptive equalization," *Proceedings IEEE*, vol. 53, pp. 1349–1387, Sept. 1985.

[3] F. R. Magee and J. G. Proakis, "Adaptive maximum-likelihood sequence estimation for digital signaling in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 19, pp. 120–124, Jan 1973.

[4] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-Survivor Processing: a general approach to MLSE in uncertain environments ," *IEEE Transactions on Communications*, vol. 43(2/3/4), pp. 354–364, February/March/April 1995.

[5] H. Kubo, K. Murakami, and T. Fujino, "An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels ," *IEEE Transactions on Communications*, vol. 42(2/3/4), pp. 1872–1880, February/March/April 1994.

[6] N. Seshadri, "Joint Data and Channel Estimation Using Blind Trellis Search Techniques," *IEEE Transactions on Communications*, vol. 42(2/3/4), pp. 1000–1011, February/March/April 1994.

[7] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for a priori unknown channels and symbol timing," *IEEE Journal of Selected Areas in Communications*, vol. 10, pp. 579–588, Mar 1992.

[8] J. W. Modestino and V. M. Eyuboglu, "Integrated multielement receiver structures for spatially distributed interference channels," *IEEE Transactions on Information Theory*, vol. 32, pp. 195–219, March 1986.

[9] J. H. Winters, "Signal acquisition and tracking with adaptive arrays in digital mobile radio system IS-54 with flat fading," *IEEE Transactions on Vehicular Technology*, vol. 2, pp. 377–384, July 1993.

[10] T. Boros, G. Raleigh, and M. Pollack, "Adaptive space-time equalization for rapidly fading communication channels," in *Proc.GLOBECOM*, pp. 984–989, 1996.

[11] J. Liang and A. Paulraj, "Two-Stage CCI/ISI Reduction with Space-Time Processing in TDMA Cellular Networks," in *Proc. Asilomar Conference on Signals and Systems*, Nov. 1996.

[12] K. Giridhar, J. J. Shynk, A. Mathur, S. Chari, and R. P. Gooch, "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Transactions on Communications*, vol. 45, pp. 473–484, April 1997.

[13] S. R. Kadaba, S. B. Gelfand, M. P. Fitz, and R. L. Kashyap, "Soft output interference suppression in TDMA wireless communications," *Wireless Networks, Baltzer Science Publishers*, pp. 507–514, 1998. Special issue on multi-user detection in wireless communications, Eds:Ryuji Kohno, Sergio Verdu and Zoran Zvonar.

[14] P. A. Ranta and A. Lappetelainen, "Application of dominant interference cancellation in street microcells," in *Proc.International Conference on Communications*, pp. 964–968, 1997.

[15] S. W. Wales, "Technique for cochannel interference suppression in TDMA mobile radio systems," *IEE Proceedings on Communications*, vol. 142, pp. 106–114, April 1995.

[16] K. Molnar and G. E. Bottomley, "D-AMPS performance in PCS bands with array processing," in *Proc.IEEE Vehicular Technology Conference*, pp. 1496–1500, May 1996.

[17] G. E. Bottomley and K. Molnar, "Interference cancellation for improved channel estimation in array processing MLSE receivers," in *Proc.IEEE Vehicular Technology Conference*, pp. 140–144, May 1997.

[18] S. N. Diggavi and A. Paulraj, "Performance of multisensor adaptive MLSE in fading channels," in *Proc.IEEE Vehicular Technology Conference*, pp. 2148–2152, May 1997.

[19] B. C. Ng, M. Cedervall, and A. Paulraj, "A structured channel estimator for maximum likelihood sequence detection," *IEEE Communication Letters*, vol. 1, pp. 52–55, March 1997.

[20] Z. Ding, "Multipath channel identification based on partial system information," *IEEE Transactions on Signal Processing*, vol. 45, pp. 235–240, Jan. 1997.

[21] S. V. Schell, D. L. Smith, and S. Roy, "Blind channel identification using subchannel response matching," in *28th Annual Conf. of Info. Science and Systems*, (Princeton, New Jersey), pp. 859–862, Mar. 1994.

[22] B. C. Ng, S. N. Diggavi, and A. Paulraj, "Joint structured channel and data estimation over time-varying channels," in *Proc. IEEE GLOBECOM*, pp. 409–413, 1997.

[23] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Transactions on Communications*, vol. 37, pp. 428–436, May 1989.

[24] T. Kailath, "Channel characterization: time-variant dispersive channels," in *Lectures on Communication System Theory* (E. J. Baghdady, ed.), pp. 95–123, New York: McGraw-Hill Book Co., 1961.

[25] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 2nd ed., 1991.

[26] L. Ljung and T. Soderstrom, *Theory and practice of recursive identification*. Cambridge, Mass.: MIT Press, 1983.

[27] K. V. Mardia and R. J. Marshall, "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, vol. 71, no. 1, pp. 135–146, 1984.

[28] A. Lapidoth, *Mismatched decoding of the multiple-access channel and some related issues in lossy source compression*. PhD thesis, Stanford University, Stanford, CA., 1995.

[29] J. G. Proakis, *Digital Communications*. New York: McGraw Hill, 2nd ed., 1989.

[30] J. H. Winters, J. Salz, and R. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Transactions on Communications*, vol. 42, pp. 1740–1751, Feb/Mar/Apr 1994.

[31] G. Caire, G. Taricco, J. Ventura-Traveset, and E. Biglieri, "A multiuser approach to narrowband cellular communications," *IEEE Transactions on Information Theory*, vol. 43, pp. 1503–1517, September 1997.

[32] W. H. Sheen and G. L. Stuber, "Mlse equalization and decoding for multipath-fading channels," *IEEE Trans. on Communications*, vol. 39, pp. 1455–1464, October 1991.

[33] J. K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulations in Rayleigh-fading channels," *IEEE Transactions on Communications*, vol. 40, pp. 74–83, January 1992.

[34] C. Schlegel, "Trellis coded modulation on time-selsective fading channels," *IEEE Transactions on Communications*, vol. 42, pp. 1617–1627, Feb/Mar/April 1994.

[35] M. J. Barrett, "Error Probability for Optimal and Suboptimal Quadratic Receivers in Rapid Rayleigh Fading Channels," *IEEE Journal on Selected Areas in Communications*, vol. SAC, no. 2, pp. 302–304, 1987.

[36] J. G. Proakis, *Digital Communications*. New York: McGraw Hill, 3 ed., 1995.

[37] B. C. Ng, *Structured channel methods in wireless communications*. PhD thesis, Stanford University, Stanford, CA., 1998.

[38] E. Eleftheriou and D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. ASSP*, vol. ASSP-34, pp. 1097–1110, Oct. 1986.

[39] J. W. Brewer, "Kronecker products and Matrix Calculus in System Theory," *IEEE Transactions on Circuits and systems*, vol. 25, pp. 772–781, September 1978.