MIMO Channel Estimation and Tracking Based on Polynomial Prediction With Application to Equalization

Yau Hee Kho, *Student Member, IEEE*, and Desmond P. Taylor, *Life Fellow, IEEE*

***Abstract*—This paper presents a multiple-input–multiple-output (MIMO) receiver design with integrated channel estimation and tracking for a time-varying frequency-selective Rician or Rayleigh fading environment. It first extends a polynomial-predictor-based channel estimation and tracking approach to a MIMO system. The structure and complexity of the estimator are similar to that of an optimum estimator using a Kalman filter, but it does not require *a priori* knowledge of the channel statistics. It employs a fixed-state transition matrix using precomputed polynomial coefficients and can be used in a Rician fading environment without reconfiguration. It is integrated with a MIMO minimum-mean-squared-error decision feedback equalizer, and simulation results show that the system performance using the estimator can be made comparable to that employing a Kalman estimator under a broad range of channel conditions.**

***Index Terms*—Equalizers, estimation, fading channels, mobile communication, multiple-input–multiple-output (MIMO) systems.**

I. Introduction

MULTIPLE-INPUT–multiple-output (MIMO) communication systems with T transmit and R receive antennas can linearly increase the available system capacity as $\min\{R, T\}$ [1], [2]. Several MIMO architectures have been proposed for the flat-fading environment. These include layered space–time (BLAST) architectures [3], [4] that transmit parallel streams of independent data to multiple receive antennas (spatial multiplexing) and space–time-coded systems [5], [6] that are designed to achieve better error rate performance through diversity and/or coding gain.

To support higher data rates, these architectures need to be extended to the wideband environment, which usually exhibits frequency-selective fading. This causes intersymbol interference (ISI) that must be compensated or equalized, and this requires accurate estimates of the channel responses. This channel estimator must estimate multiple parameters and track temporal variations of the channel.

Manuscript received February 8, 2007; revised June 25, 2007, August 15, 2007, and August 19, 2007. The work of Y. H. Kho was supported in part by the University of Canterbury Doctoral Scholarship and in part by the Institution of Engineering and Technology (IET, formerly the IEE) Hudswell International Research Scholarship. The review of this paper was coordinated by Prof. W. Su.

The authors are with the Department of Electrical and Computer Engineering, University of Canterbury, Christchurch 8140, New Zealand (e-mail: yhkho@ieee.org; taylor@elec.canterbury.ac.nz).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.
Digital Object Identifier 10.1109/TVT.2007.907318

Statistics-based methods such as the Kalman filter, which assumes a low-order autoregressive (AR) model [7] of the channel state, produce excellent estimates as they incorporate a dynamic model of the channel into their operation. A Kalman filter produces a minimum variance estimate of the parameters being tracked [7]. However, it requires *a priori* knowledge of the channel and noise statistics. Their acquisition usually requires a long measurement time [8] that adds complexity to the estimation process. In [9], a Kalman-based channel estimator is used to estimate and track the frequency-selective channel responses. The AR parameters needed by the Kalman filter and the noise statistics are assumed known. Additional algorithms to obtain the noise statistics can be used [10], where a noise covariance estimation algorithm and a noise whiteness test are developed to estimate the noise covariance needed by the Kalman filter.

It is desirable that an estimator be able to achieve Kalman-like performance without requiring *a priori* knowledge of the channel and noise statistics. Motivated by this, we develop a MIMO channel estimator that is comparable in complexity and performance to a Kalman estimator but does not require *a priori* knowledge of the channel and noise statistics. Moreover, it can be used in a Rician fading environment without reconfiguration of the state transition matrix to accommodate the nonrandom mean components of the channel responses. The algorithm is a vector extension of the polynomial-based generalized recursive least squares (GRLS) algorithm [11]. Due to the vector received signal, we call the proposed channel estimator a polynomial-predictor-based vector GRLS (VGRLS) estimator. We evaluate its steady-state performance with known training symbols under various system and channel settings. We then evaluate the performance of an integrated receiver using the estimator and a MIMO minimum-mean-squared-error (MMSE) decision feedback equalizer (DFE) similar to that in [9].

We describe the signal model in Section II. In Section III, we develop the channel estimator, and in Section IV, we describe the example receiver. Simulation results and discussions are presented in Section V, and finally, conclusions are presented in Section VI.

II. Signal and Channel Models

A. General Model

We assume a MIMO system transmitting independent signals from each of the T antennas to $R \geq T$ receive antennas. Fig. 1

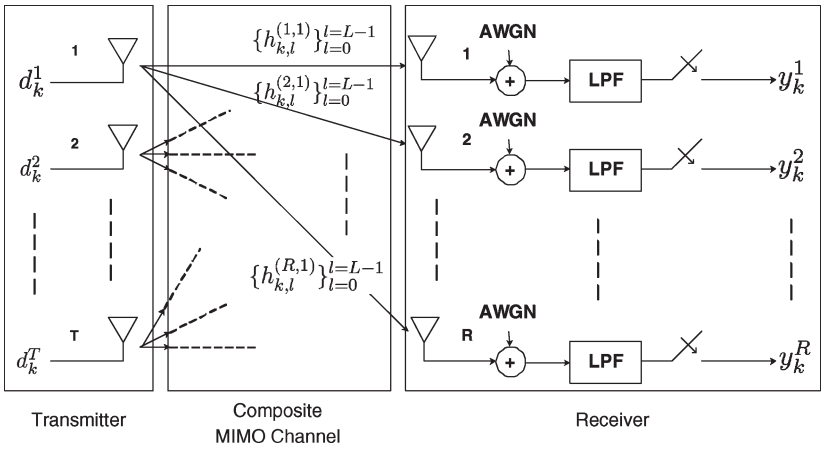


Fig. 1. General block diagram of a discrete-time MIMO communication system at time k for T transmit and R receive antennas.

shows a discrete-time model for this (T, R) MIMO system. At the receiver, each of the R antennas receives a linear combination of the transmitted signals. The jth symbol-rate-sampled representation of the complex baseband received signal at time k may be written as

$$y_k^{(j)} = \sum_{i=1}^{T} \sum_{l=0}^{L-1} d_{k-l}^{(i)} h_{k,l}^{(j,i)} + n_k^{(j)}, \qquad j = 1, 2, \ldots, R \quad (1)$$

where $d_k^{(i)}$ is the kth transmitted complex baseband M-ary data symbol from the ith antenna, $\{h_{k,l}^{(j,i)}\}_{l=0}^{l=L-1}$ is the sampled fading dispersive composite channel impulse response (convolution of the transmit pulse shape and physical channel response) between the ith transmit and jth receive antennas at time k with delay spread of L symbol periods, and $n_k^{(j)}$ is the sampled additive white Gaussian noise (AWGN) at the jth receive antenna with variance σ_n^2.

The MIMO received signal of (1) may be expressed in matrix–vector form [12] as

$$\mathbf{y}_k = \sum_{l=0}^{L-1} \mathbf{H}_{k,l} \mathfrak{d}_{k-l} + \mathbf{n}_k \quad (2)$$

where we define the vectors

$$\mathbf{y}_k = \begin{pmatrix} y_k^{(1)} \\ y_k^{(2)} \\ \vdots \\ y_k^{(R)} \end{pmatrix} \qquad \mathfrak{d}_k = \begin{pmatrix} d_k^{(1)} \\ d_k^{(2)} \\ \vdots \\ d_k^{(T)} \end{pmatrix} \qquad \mathbf{n}_k = \begin{pmatrix} n_k^{(1)} \\ n_k^{(2)} \\ \vdots \\ n_k^{(R)} \end{pmatrix} \quad (3)$$

and the $R \times T$ channel matrix taps

$$\mathbf{H}_{k,l} = \begin{pmatrix} h_{k,l}^{(1,1)} & \cdots & h_{k,l}^{(1,T)} \\ \vdots & \ddots & \vdots \\ h_{k,l}^{(R,1)} & \cdots & h_{k,l}^{(R,T)} \end{pmatrix}, \qquad l = 0, 1, 2, \ldots, L-1. \quad (4)$$

To derive the VGRLS estimator from the GRLS estimator in [11], we reformulate (2). We observe that there are L channel matrix taps. We represent each as a column vector using the operator $\mathrm{vec}(\mathbf{H}_{k,l})$ and stack the columns of $\mathbf{H}_k = [\mathbf{H}_{k,0}, \ldots, \mathbf{H}_{k,L-1}]$ into a single-length RTL channel vector

$$\begin{aligned} \mathfrak{h}_k &= \mathrm{vec}(\mathbf{H}_k) \\ &= \Big[h_{k,0}^{(1,1)} \cdots h_{k,0}^{(R,1)} \cdots h_{k,0}^{(1,T)} \cdots h_{k,0}^{(R,T)} \\ &\qquad \ldots, h_{k,L-1}^{(1,1)} \cdots h_{k,L-1}^{(R,1)} \cdots h_{k,L-1}^{(1,T)} \cdots h_{k,L-1}^{(R,T)}\Big]^t \end{aligned} \quad (5)$$

where t denotes matrix transposition. To ensure dimensional compatibility, we define an $R \times RTL$ transmitted data matrix $\mathfrak{D}_k$ from (3) as

$$\begin{aligned} \mathfrak{D}_k = \Big[&d_k^{(1)} \cdots d_k^{(T)}, d_{k-1}^{(1)}, \ldots, d_{k-1}^{(T)} \\ &\ldots, d_{k-L+1}^{(1)} \cdots d_{k-L+1}^{(T)}\Big] \otimes \mathbf{I}_R \end{aligned} \quad (6)$$

where $\mathbf{I}_R$ is the $R \times R$ identity matrix, and $\otimes$ is the Kronecker product. We may then write (2) in the form of

$$\mathbf{y}_k = \mathfrak{D}_k \mathfrak{h}_k + \mathbf{n}_k. \quad (7)$$

B. Channel Model

The RT MIMO subchannels are each assumed to exhibit wide sense stationary uncorrelated scattering (WSSUS) [13]. We assume that each of the coefficients $h_{k,l}^{(j,i)}$ evolves according to Clarke's fading model [14] under common fading conditions. We also assume that each channel coefficient consists of a nonrandom (specular) component and a random (diffuse) component as $h_{k,l}^{(j,i)} = h_l^{(nr),(j,i)} + h_{k,l}^{(r),(j,i)}$. The power ratio between the specular and diffuse components is given by the Rice K-factor

$$K = \frac{\left|h^{(nr)}\right|^2}{E\left\{\left|h^{(r)}\right|^2\right\}} \quad (8)$$

where a K value of 0 corresponds to Rayleigh fading, and a large K corresponds to Rician fading. In reality, a specular component can be present in any or all of the paths, and the value of the K-factor can be the same or different for

each path. For simplicity, we assume here that all the multipath components contain a specular component with the same value of K.

III. Channel Estimation

A. Statistical State-Space Model

Channel estimators based on the Kalman filter [9], [15] typically assume that the multipath fading channel response vector (5) evolves according to an order P_a vector AR (VAR) process [16], which may be modeled by the state equation

$$\mathbf{h}_{k+1} = \mathbf{A}\mathbf{h}_k + \mathbf{v}_k \tag{9}$$

where

$$\mathbf{h}_k = \left[\mathfrak{h}_k^t, \mathbf{h}_{k-1}^t, \ldots, \mathbf{h}_{k-P_a+1}^t\right]^t \tag{10}$$

is the $RTLP_a \times 1$ channel state vector at time k consisting of P_a "time-shifted" vectors of (5), and $\mathbf{v}_k$ is the kth zero-mean process noise vector of dimension $RTLP_a \times 1$ such that

$$E\left\{\mathbf{v}_k \mathbf{v}_l^H\right\} = \begin{cases} \mathbf{R}_v, & \text{for } k = l \\ \mathbf{0}_{m,m}, & \text{for } k \neq l \end{cases} \tag{11}$$

with $\mathbf{0}_{m,m}$ being the $(m \times m)$ null matrix, and $m = RTLP_a$. The superscript H denotes Hermitian transposition, and $\mathbf{A}$ is the $RTLP_a \times RTLP_a$ state transition matrix having the form

$$\mathbf{A} = \begin{pmatrix} \mathbb{A}_1 & \mathbb{A}_2 & \cdots & \mathbb{A}_{P_a-1} & \mathbb{A}_{P_a} \\ & \mathbf{I}_{RTL(P_a-1)} & & & \mathbf{0}_{RTL(P_a-1),RTL} \end{pmatrix} \tag{12}$$

where the matrices $\{\mathbb{A}_l\}, l = 1, 2, \ldots, P_a$ are the $RTL \times RTL$ matrix coefficients of the VAR process. These and the process noise autocovariance matrix $\mathbf{R}_v$ may be obtained by measuring the channel statistics and solving the resulting matrix–vector Yule–Walker equations [16]. The choice of the process order P_a is a tradeoff between complexity and modeling accuracy [15]. When a high degree of accuracy is needed, a large P_a is selected such that the variances of the elements of $\mathbf{v}_k$ are small [11]. We may express the MIMO received signal of (7) as

$$\mathbf{y}_k = \mathbf{d}_k \mathbf{h}_k + \mathbf{n}_k \tag{13}$$

where the $R \times RTLP_a$ data matrix $\mathbf{d}_k$ is defined as

$$\mathbf{d}_k = \begin{bmatrix} \mathfrak{D}_k & | & \mathbf{0}_{R,RTL(P_a-1)} \end{bmatrix} \tag{14}$$

with $\mathfrak{D}_k$ given by (6). The state-space model used by the Kalman estimator [7], [9], [15] is given by (9) and (13). As structured, it is restricted to Rayleigh fading channels. However, it may be explicitly modified to model specular components [15] by restructuring the state transition matrix $\mathbf{A}$.

B. Polynomial-Based State-Space Model

The state equation of (9) can be interpreted as a one-step length-P_a vector–matrix predictor of the channel state vector with the VAR matrix coefficients $\{\mathbb{A}_l\}$ for $l = 1, 2, \ldots, P_a$ being the one-step prediction coefficients and $\mathbf{v}_k$ the associated prediction error.

TABLE I
Polynomial Coefficients of Various Order and Length

Length P	*Order N*	*Polynomial Coefficients* $\{a_1, a_2, \cdots, a_P\}$
2	0	{1/2, 1/2}
2	1	{2, -1}
3	0	{1/3, 1/3, 1/3}
3	1	{4/3, 1/3, -2/3}
3	2	{3, -3, 1}
4	0	{1/4, 1/4, 1/4, 1/4}
4	1	{1, -1/2, 0, 1/2}
4	2	{9/4, -3/4, -5/4, 3/4}
4	3	{4, -6, 4, -1}

If we assume the fading processes to vary smoothly, we may model the time evolution of each of their samples as polynomial sequences of order N [11] based on truncated Taylor series (polynomial) expansions of the fading process [13] in the time domain[1] over a sufficiently small window. From the theory of polynomial prediction [17], a one-step predictor of length P with coefficients $\{a_p\}$ for $p = 1, 2, \ldots, P$ may be derived for each polynomial sequence. Following [17]–[19], for the μth scalar channel sample in (5) for $\mu = 1, 2, \ldots, RTL$, we may write a one-step Nth-order polynomial prediction equation at time k as

$$h_{k,\mu} = \sum_{p=1}^{P} a_p h_{k-p,\mu} + e_{k,\mu}(N, P) \tag{15}$$

where P is the length of the polynomial predictor assuming that each of the channel coefficient is modeled as a truncated t power series [13] of order N and that the series converges over a window of size $P + 1$ [20].

The polynomial predictor coefficients $\{a_p\}$ for $p = 1, 2, \ldots, P$ are dependent only on the values of N and P, where $P \geq N + 1$, and may be computed using a Lagrange multiplier technique [18] or a standard least-square optimization approach [11]. Note that the computation does not require any channel statistics. The polynomial coefficients for various orders N and lengths P are given in Table I, as calculated in [11]. We note that the norm of the coefficients gets larger as N and P increase, and this tends to degrade the performance of the estimator at low SNR, as will be shown later.

The prediction error arising from the truncation of the series to the first N terms $e_{k,\mu}(N, P)$ is dependent on the order of the polynomial series and the predictor length, where $e_{k,\mu}(N, P) \to 0$ as $N \to \infty$ [11]. It will be small if the window of expansion (i.e., the predictor length, P) is small, thereby allowing the use of a small value of N. As will be shown later, a larger value of P is only required in very fast fading. Using (15), a VAR-like model of the channel vector in (5) may be formulated as

$$\mathfrak{h}_k = \sum_{p=1}^{P} \mathfrak{U}_p \mathfrak{h}_{k-p} + \mathbf{e}_k(N, P) \tag{16}$$

[1] In [13], these are known as t-power series expansions.

where the $RTL \times RTL$ polynomial predictor matrices are given by $\mathfrak{U}_p = a_p \mathbf{I}_{RTL}$ for $p = 1, 2, \ldots, P$. The model is only VAR like because the error vector

$$\mathbf{e}_k(N,P) = \Big[e_{k,0}^{(1,1)} \cdots e_{k,0}^{(R,1)} \cdots e_{k,0}^{(1,T)} \cdots e_{k,0}^{(R,T)} \cdots e_{k,L-1}^{(1,1)} \cdots e_{k,L-1}^{(R,1)} \cdots e_{k,L-1}^{(1,T)} \cdots e_{k,L-1}^{(R,T)}\Big]^t \quad (17)$$

is not necessarily zero mean or white [11], as required by a VAR process. In general, the elements of the covariance of $\mathbf{e}_k(N,P)$ will be small over a suitably small window of expansion around each sampling instant. As a result, if $\mathbf{e}_k(N,P)$ is assumed to be zero, a state-space model similar in form to (9) but with unforced dynamics may be approximately formulated from (16) as

$$\mathbf{h}_{k+1} = \mathbf{U}\mathbf{h}_k \quad (18)$$

where

$$\mathbf{h}_k = \left[\mathfrak{h}_k^t, \mathfrak{h}_{k-1}^t, \ldots, \mathfrak{h}_{k-P+1}^t\right]^t \quad (19)$$

is the $RTLP \times 1$ channel state vector at time k, and the associated state transition matrix is given by

$$\mathbf{U} = \begin{pmatrix} \mathfrak{U}_1 & \mathfrak{U}_2 & \cdots & \mathfrak{U}_{P-1} & \mathfrak{U}_P \\ & \mathbf{I}_{RTL(P-1)} & & & \mathbf{0}_{RTL(P-1),RTL} \end{pmatrix}. \quad (20)$$

This matrix is similar in form to (12) with P_a replaced by P and the matrices $\mathbb{A}_l$ replaced by the matrices $\mathfrak{U}_p$. The observation equation associated with (18) is similar to (13), except that P_a becomes P.

Equations (13) and (18) define a nonstatistical polynomial-based state-space model with unforced dynamics. It does not require channel statistics in the derivation of the state transition matrix coefficients and can be used with both Rayleigh and Rician fading channels with no explicit reconfiguration of the state transition matrix $\mathbf{U}$.

C. Channel Estimator

In [21], an estimator using the Kalman recursive least squares (RLS) algorithm that incorporates a state-space model in the process was developed. The conventional RLS algorithm is a special case of this GRLS algorithm. It models each of the sample of the channel responses as a two-term t-power series [13]. Instead of estimating the channel response coefficients, the time-invariant coefficients of the t-power series are estimated. As a two-term t-power series is only suitable for linearly time-varying channels [18], the resulting state-space model is limited to channels that linearly vary with time.

Here, we employ the state-space model of (18) in a vector-based GRLS algorithm to directly estimate the channel tap or state vector $\mathbf{h}_k$. The coefficients of the state transition matrix (20) are predetermined (cf. Table I) for a given predictor length P and polynomial order N.

Assuming that $\hat{\mathbf{h}}_{k/k-1}$ and $\mathbf{P}_{k/k-1}$ are known, the update equations for the VGRLS algorithm may be written following [11] as

$$\mathbf{K}_k = \mathbf{P}_{k/k-1}\mathbf{d}_k^H \left(\mathbf{I}_R + \mathbf{d}_k \mathbf{P}_{k/k-1}\mathbf{d}_k^H\right)^{-1} \quad (21)$$

$$\mathbf{P}_{k/k} = (\mathbf{I}_{RTLP} - \mathbf{K}_k\mathbf{d}_k)\mathbf{P}_{k/k-1} \quad (22)$$

$$\hat{\mathbf{h}}_{k/k} = \hat{\mathbf{h}}_{k/k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{d}_k\hat{\mathbf{h}}_{k/k-1}). \quad (23)$$

The prediction equations may then be written as

$$\hat{\mathbf{h}}_{k+1/k} = \mathbf{U}\hat{\mathbf{h}}_{k/k} \quad (24)$$

$$\mathbf{P}_{k+1/k} = \lambda^{-1}\mathbf{U}\mathbf{P}_{k/k}\mathbf{U}^H \quad (25)$$

where $\hat{\mathbf{h}}_{k/k-1}$ is the estimate of the channel state vector at time k based on $(k-1)$ prior received samples, λ is the RLS "forget factor," $\mathbf{K}_k$ is analogous to the Kalman gain vector [7], and $\mathbf{P}_{k/k}$ is the so-called "intermediate" matrix.[2] To initialize the algorithm, we set the estimated channel state vector $\hat{\mathbf{h}}_{1/0}$ to the null vector and let $\mathbf{P}_{1/0} = \delta^{-1}\mathbf{I}_{LP}$, where δ is a small positive real constant. Note that when $P = 1$ and $N = 0$, the VGRLS algorithm reduces to a conventional vector RLS estimation algorithm [11] whose tracking performance can be estimated [22].

We note that the VGRLS algorithm of (21)–(25) is similar in structure to a Kalman filter as both consist of time-update and prediction equations. Due to the Riccatti recursion in (22), the complexity of VGRLS in the highest term is $\mathcal{O}((RTLP)^3)$, which is similar to that of the Kalman filter. Thus, the "baseline" complexity of the two algorithms is similar. However, the VGRLS does not require any channel statistics to derive the coefficients of the state transition matrix.

IV. Vector DFE Receiver

As an example of the application of the VGRLS estimator, we integrate it into a receiver that employs a vector DFE structure[3] similar to those in [9] and [23]. For the ISI-corrupted received signals, maximum likelihood sequence estimation is the optimum equalization method [24], [25]. However, for a given modulation size, its complexity exponentially increases with the channel delay spread L and the number of transmit antennas T. Furthermore, its decision delay is significant (typically about $5L$) [24]. We have therefore employed an MMSE DFE [23] structure.

The resulting receiver structure is shown in Fig. 2, where the channel estimator provides vector estimates $\{\hat{h}_{k,l}^{(j,i)}\}_{l=0}^{l=L-1}$ of the channel responses $\{h_{k,l}^{(j,i)}\}_{l=0}^{l=L-1}$ for the adaptive equalization of the received signal streams $\{y_k^{(j)}\}$ for $j = 1, 2, \ldots, R$ and $i = 1, 2, \ldots, T$. Initially, the receiver operates in training mode, where only the estimator is operating, and a training sequence is used to obtain an initial channel estimate. Following this, the receiver operates in a decision-directed mode, where the estimator and equalizer work in tandem.

[2] $\mathrm{P}_{k/k}$ is the inverse input autocorrelation matrix in a conventional RLS algorithm.

[3] For the vector DFE in this paper, the channel estimate is directly used to calculate the equalizer tap coefficients.

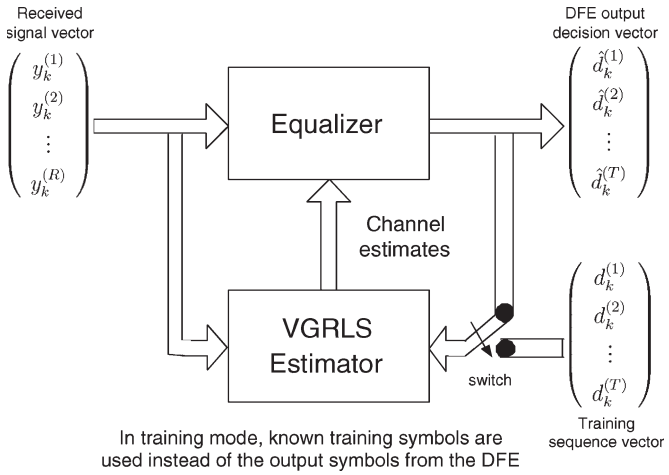


Fig. 2. Overall receiver structure in a decision-directed mode, where the vector DFE and VGRLS estimator work in tandem. Note initially that when in training mode, the estimator operates alone using the known training symbols instead of the output symbols from the DFE.

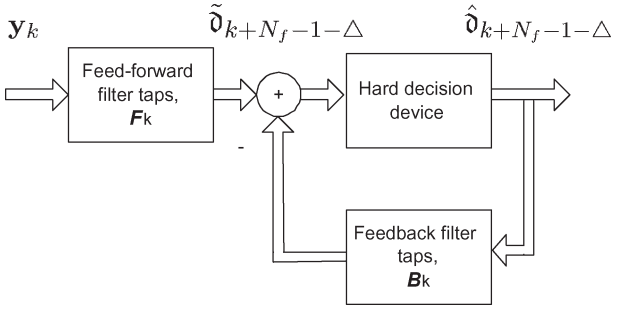


Fig. 3. Vector MMSE DFE.

We assume that the DFE contains N_f feedforward filter matrix taps ($\mathbf{F}_k$) and N_b feedback filter matrix taps ($\mathbf{B}_k$), as shown in Fig. 3. These matrix tap coefficients are jointly optimized based on the MMSE performance criterion. The design of the optimum MMSE vector DFE is described in detail in [23].

At each time k, the feedforward filter processes a block of N_f received signal vectors $\mathbf{y}_k$ of (2), which can be written in matrix form in (26), shown at the bottom of the next page.

This may be expressed in compact form as

$$\mathbf{y}_{k+N_f-1:k} = \mathbf{C}\mathfrak{d}_{k+N_f-1:k-L+1} + \mathbf{n}_{k+N_f-1:k} \tag{27}$$

where $\mathbf{C}$ is the convolution matrix given in (28), shown at the bottom of the next page.

Corresponding to this is the block of $(N_f + L - 1)$ input symbol vectors $\mathfrak{d}_{k+N_f-1:k-L+1}$ consisting of $(L-1)$ past symbol vectors $\mathfrak{d}_{k-1:k-L+1}$ and $(N_f - 1)$ future symbol vectors $\mathfrak{d}_{k+N_f-1:k+1}$ that have yet to be detected. The feedback filter utilizes a subset $\mathfrak{d}_{k-1:k-N_b}$ of previously detected symbol vectors to cancel their interfering effect on the current symbol vector $\mathfrak{d}_k$.

We define the $T(N_f + L - 1) \times T(N_f + L - 1)$ input autocorrelation matrix

$$\mathbf{R}_{\mathfrak{d}\mathfrak{d}} = E\left\{\mathfrak{d}_{k+N_f-1:k-L+1}\mathfrak{d}^H_{k+N_f-1:k-L+1}\right\} \tag{29}$$

and the $(RN_f) \times (RN_f)$ noise autocorrelation matrix

$$\mathbf{R}_{nn} = E\left\{\mathbf{n}_{k+N_f-1:k}\mathbf{n}^H_{k+N_f-1:k}\right\}. \tag{30}$$

The input–output cross correlation and output autocorrelation matrices then follow as

$$\mathbf{R}_{\mathfrak{d}y} = E\left\{\mathfrak{d}_{k+N_f-1:k-L+1}\mathbf{y}^H_{k+N_f-1:k}\right\} = \mathbf{R}_{\mathfrak{d}\mathfrak{d}}\mathbf{C}^H \tag{31}$$

$$\begin{aligned}\mathbf{R}_{yy} &= E\left\{\mathbf{y}_{k+N_f-1:k}\mathbf{y}^H_{k+N_f-1:k}\right\} \\ &= \mathbf{C}\mathbf{R}_{\mathfrak{d}\mathfrak{d}}\mathbf{C}^H + \mathbf{R}_{nn}.\end{aligned} \tag{32}$$

The vector DFE consists of a feedforward filter matrix

$$\mathbf{F}^H_k = \left[\,\mathbf{F}^H_{k,0} \quad \mathbf{F}^H_{k,1} \quad \cdots \quad \mathbf{F}^H_{k,N_f-1}\,\right] \tag{33}$$

and a feedback filter matrix

$$\mathbf{B}^H_k = \left[\,\mathbf{B}^H_{k,1} \quad \cdots \quad \mathbf{B}^H_{k,N_b}\,\right]. \tag{34}$$

For analytical convenience, we define an *extended* $T \times T(N_f + L - 1)$ matrix $\tilde{\mathbf{B}}^H_k = [\,\mathbf{0}_{T,T\triangle} \quad \mathbf{I}_T \quad \mathbf{B}^H_k\,]$ that corresponds to the symbol vectors $\mathfrak{d}_{k+N_f-1:k-L+1}$ in (27). Note that $\triangle$ is a decision delay that satisfies the condition $(\triangle + N_b + 1) = (N_f + L - 1)$. In general, for ISI cancellation, we require $N_b \geq L - 1$. For the purpose of modeling, we assume here that $N_b = L - 1$ so that $\triangle = N_f - 1$ [23].

The vector DFE's error vector at time k, assuming correct past decisions, is given by

$$\begin{aligned}\mathbf{E}_k &= \mathfrak{d}_{k+N_f-1-\triangle} - \tilde{\mathfrak{d}}_{k+N_f-1-\triangle} \\ &= \mathfrak{d}_{k+N_f-1-\triangle} - \sum_{f=0}^{N_f-1}\mathbf{F}^H_{k,f}\mathbf{y}_{k+N_f-1-f} \\ &\quad + \sum_{b=1}^{N_b}\mathbf{B}^H_{k,b}\mathfrak{d}_{k+N_f-1-\triangle-b} \quad (\text{putting } \triangle = N_f - 1) \\ &= \left[\,\mathbf{0}_{T,T\triangle} \quad \mathbf{I}_{T,T} \quad \mathbf{B}^H_{k,1}\cdots \quad \mathbf{B}^H_{k,N_b}\,\right]\mathfrak{d}_{k+N_f-1:k-L+1} \\ &\quad - \left[\,\mathbf{F}^H_{k,0} \quad \cdots \quad \mathbf{F}^H_{k,N_f-1}\,\right]\mathbf{y}_{k+N_f-1:k} \\ &= \tilde{\mathbf{B}}^H_k\mathfrak{d}_{k+N_f-1:k-L+1} - \mathbf{F}^H_k\mathbf{y}_{k+N_f-1:k}.\end{aligned} \tag{35}$$

The error autocorrelation matrix may then be written as

$$\mathbf{R}_{ee} = E\left[\mathbf{E}^H_k\mathbf{E}_k\right] = \tilde{\mathbf{B}}^H_k\mathbf{R}^{-1}\tilde{\mathbf{B}}_k \tag{36}$$

where

$$\mathbf{R} = \mathbf{R}^{-1}_{\mathfrak{d}\mathfrak{d}} + \mathbf{C}^H\mathbf{R}^{-1}_{nn}\mathbf{C}. \tag{37}$$

To calculate the feedback taps, we partition $\mathbf{R}$ into the submatrix form [23]

$$\mathbf{R} = \begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}^H_{12} & \mathbf{R}_{22}\end{pmatrix} \tag{38}$$

where $\mathbf{R}_{11}$ is the $T(\triangle+1)\times T(\triangle+1)$ upper left submatrix. We further define a matrix

$$\mathbf{G}^t = [\,\mathbf{0}_{T,T\triangle} \quad \mathbf{I}_T\,] \tag{39}$$

and from (38) and (39), we obtain

$$\widetilde{\mathbf{B}}_k = \begin{bmatrix} \mathbf{I}_{T(\triangle+1)} \\ \mathbf{R}_{12}^H\mathbf{R}_{11}^{-1} \end{bmatrix} \mathbf{G} = \begin{bmatrix} \mathbf{0}_{T,T\triangle} \\ \mathbf{I}_T \\ \mathbf{B}_k \end{bmatrix} \tag{40}$$

where $\widetilde{\mathbf{B}}_\mathbf{k}$ is the *extended* feedback matrix containing $\mathbf{B}_k$ as the optimal feedback matrix tap coefficients. The error autocorrelation matrix of (36) can then be calculated and the optimal $\triangle$ determined such that the trace of $\mathbf{R}_{ee}$ is minimized [23]. As we have assumed that the number of feedback matrix taps is $N_b = L-1$, the delay $\triangle$ is fixed at N_f-1, which is found to be optimal for most practical channels [23]. The MMSE feedforward matrix tap coefficients are calculated as

$$\mathbf{F}_k^H = \widetilde{\mathbf{B}}_k^H \mathbf{R}_{\mathfrak{d}y}\mathbf{R}_{yy}^{-1}. \tag{41}$$

At each time instant k where we formulate the $RN_f \times T(N_f+L-1)$ block prewindowed channel convolution matrix $\hat{\mathbf{C}}$, the estimates of (28), where $\hat{\mathbf{H}}_{b,m}$ for $b=\{k+N_f-1,\ldots,k\}$ and $m=\{0,1,\ldots,L-1\}$, are the estimates of the $(R\times T)$ channel matrices $\mathbf{H}_{k,l}$ of (4). Using the estimate $\hat{\mathbf{C}}$ in place of $\mathbf{C}$, the matrix tap coefficients of the DFE are then estimated following the steps described above [23].

The receiver operates in two modes.

A. Training Mode

In this mode, only the VGRLS estimator is operating using a training sequence of length L_t according to the following.

Step 1) Initiate the VGRLS algorithm with an all-zero estimated channel vector $\hat{\mathbf{h}}_{1/0}$ and an "intermediate" matrix $\mathbf{P}_{1/0} = \delta^{-1}\mathbf{I}_{RTLP}$, where δ is a small positive real constant, and $\mathbf{I}_{RTLP}$ is an identity matrix with dimension of $RTLP \times RTLP$. Using the observation vector $\mathbf{y}_k$, compute the Kalman gain (21), update the "intermediate" matrix (22), and update the estimated channel vector (23).

Step 2) Compute the one-step predicted channel vector (24) and the one-step predicted "intermediate" matrix (25).

Step 3) With every subsequent received observation vector $\mathbf{y}_k$, $k\geq 2$, repeat Steps 2) and 3) until the end of the training sequence.

B. Decision-Directed Mode

In this mode, the VGRLS estimator and the DFE operate together. Due to the equalizer decision delay of $\triangle = N_f-1$ symbols with reference to the equalizer input, a time lag is introduced, where at time $k-1$, the output symbols from the DFE are delayed by $\triangle$ symbol periods. Thus, the output of the decision device is the estimated symbol vector [4] $\hat{\mathfrak{d}}_{k-\triangle-1} = \{\hat{d}^{(1)}_{k-\triangle-1}, \hat{d}^{(2)}_{k-\triangle-1}, \ldots, \hat{d}^{(T)}_{k-\triangle-1}\}^t$. This is fed to the VGRLS estimator in place of the training symbols to provide the next channel estimate vector at time k. Using the DFE decision vectors $\hat{\mathfrak{d}}_{k-\triangle-1},\ldots,\hat{\mathfrak{d}}_{k-\triangle-L}$, the received vector $\mathbf{y}_{k-\triangle-1}$, and the P previously estimated channel vectors, the VGRLS produces $\hat{\mathbf{h}}_{k-\triangle}$.[5] To calculate the vector DFE at time k, the N_f most recent estimated channel vectors are needed. Up to time $k-\triangle$, the channel estimates are available from the VGRLS estimator, and the last $\triangle$ channel vectors need to be predicted. A simple method is to assume that the channel remains constant over $\triangle$ time symbols so that $\hat{\mathbf{h}}_k = \hat{\mathbf{h}}_{k-1} = \cdots = \hat{\mathbf{h}}_{k-\triangle}$, where $\hat{\mathbf{h}}_{k-\triangle}$ is available from the estimator. However, this applies only to a very slowly fading channel.

As an alternative, we employ a polynomial prediction module similar to that of [18] for predicting the $\triangle$ channel vectors. Since the underlying structure of the VGRLS estimator uses a t-power series expansion [13] for modeling the channel fading process as an Nth-order polynomial series, the polynomial-based state transition matrix of (20) is already available.

[4]For convenience of illustration, we shift the time index of the DFE in this section from $k+N_f-1:k$ to $k:k-N_f+1$ so that the output of the decision device at time k is indexed as $\hat{\mathfrak{d}}_{k-\triangle}$ instead of $\hat{\mathfrak{d}}_{k+N_f-1-\triangle}$.

[5]For brevity, we simplify the notation $\hat{\mathbf{h}}_{k-\triangle/k-\triangle-1}$ to $\hat{\mathbf{h}}_{k-\triangle}$.

$$\begin{pmatrix} \mathbf{y}_{k+N_f-1} \\ \mathbf{y}_{k+N_f-2} \\ \vdots \\ \mathbf{y}_k \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{k+N_f-1,0} & \cdots & \mathbf{H}_{k+N_f-1,L-1} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{k+N_f-2,0} & \cdots & \mathbf{H}_{k+N_f-2,L-1} & 0 & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & 0 & \mathbf{H}_{k,0} & \cdots & \mathbf{H}_{k,L-1} \end{pmatrix} \begin{pmatrix} \mathfrak{d}_{k+N_f-1} \\ \mathfrak{d}_{k+N_f-2} \\ \vdots \\ \mathfrak{d}_{k-L+1} \end{pmatrix} + \begin{pmatrix} \mathbf{n}_{k+N_f-1} \\ \mathbf{n}_{k+N_f-2} \\ \vdots \\ \mathbf{n}_k \end{pmatrix} \tag{26}$$

$$\mathbf{C} = \begin{pmatrix} \mathbf{H}_{k+N_f-1,0} & \cdots & \mathbf{H}_{k+N_f-1,L-1} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{k+N_f-2,0} & \cdots & \mathbf{H}_{k+N_f-2,L-1} & 0 & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & 0 & \mathbf{H}_{k,0} & \cdots & \mathbf{H}_{k,L-1} \end{pmatrix} \tag{28}$$

It is therefore straightforward to compute the predicted channel vectors as

$$\begin{aligned}\hat{\mathbf{h}}_{k-\triangle+1} &= \mathbf{U}\hat{\mathbf{h}}_{k-\triangle} \\ \hat{\mathbf{h}}_{k-\triangle+2} &= \mathbf{U}\hat{\mathbf{h}}_{k-\triangle+1} \\ \vdots \quad &= \quad \vdots \\ \hat{\mathbf{h}}_{k} &= \mathbf{U}\hat{\mathbf{h}}_{k-1}. \end{aligned} \tag{42}$$

This method of channel vector prediction (42) is similar in form to that used in [9].

The channel estimates provided by the VGRLS estimator and the channel prediction module are used to compute the feedforward and feedback tap coefficients of the DFE. The received signal vectors are equalized by the DFE, and a detected signal vector is produced at the output of the decision device. These are used at the input of the VGRLS estimator in decision-directed operation. The operations during decision-directed mode may be summarized as follows.

- Step 1) With $\hat{\mathbf{h}}_{k-\triangle-1}$ available at time $k-1$, operate the VGRLS estimator to produce $\hat{\mathbf{h}}_{k-\triangle}$ at time k using the DFE decisions $\hat{\mathfrak{d}}_{k-\triangle-1}, \ldots, \hat{\mathfrak{d}}_{k-\triangle-L}$, the received vector $\mathbf{y}_{k-\triangle-1}$, and the P previously estimated channel vectors.
- Step 2) Predict the next $\triangle$ channel vectors, as in (42).
- Step 3) Formulate $\hat{\mathbf{C}}$, i.e., the estimated convolution matrix of (28).
- Step 4) Calculate the optimum coefficients of the DFE matrix tap vectors $\mathbf{B}_k$ and $\mathbf{F}_k$ of (40) and (41).
- Step 5) Equalize the received vectors, and obtain $\hat{\mathfrak{d}}_{k-\triangle}$.
- Step 6) At the next time instance, repeat Steps 1)–6).

V. Simulation Results and Discussion

We now evaluate the performance of the VGRLS channel estimator and the integrated estimator and vector DFE. We assume throughout an uncoded VBLAST-type [4] MIMO system. Independent QPSK signal streams are transmitted from each transmit antenna. Each transmitted frame consists of $L_t = 26$ training symbols and $L_d = 116$ data symbols, unless stated otherwise. We assume independent WSSUS subchannels each with similar fading conditions. The fading processes are assumed to follow Clarke's model [14] and are simulated according to [26]. Each subchannel is assumed to have a delay spread of $L = 3$ symbols and to have a uniform power delay profile with three rays. Each of the multipath rays may contain both a specular and a random component. The Rice K-factor in (8) defines the power ratio between the specular and random components.

We evaluate the performance of the VGRLS estimator in terms of the "mean square deviation" (MSD), which is the squared norm difference between the actual and estimated channel responses. The estimator is operated alone and constantly updates the estimated channel responses using knowledge of the transmitted signals. It is assumed to operate in transient mode during the L_t symbol training sequence, after which, it is assumed to operate in steady-state mode. The MSD measures this steady-state performance of the estimator, where the first L_t symbols of each frame are not included in the MSD calculation. At the beginning of a new frame, the estimator reinitializes and starts the channel acquisition again. The MSD performance versus the SNR of the estimator with a fixed predictor length of $P = 3$ and 4, and various polynomial orders $N = 0, 1, 2, 3$ is evaluated. The steady-state MSD in a given αth frame is estimated as

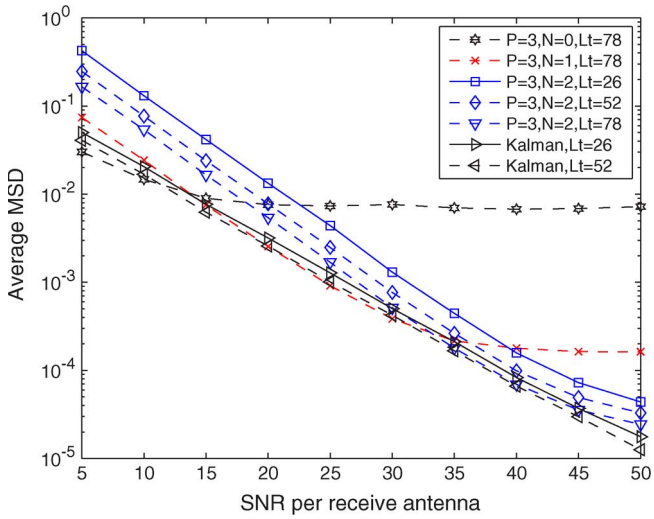


Fig. 4. MSD of the VGRLS estimator and that of a Kalman filter for a (2, 2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T$ of 0.002. VGRLS with $N = 0$ corresponds to a conventional vector RLS algorithm. With sufficient training sequence length, the MSD of VGRLS with $N = 2$ approaches that of a Kalman filter's.

$$\sigma^2_{\mathrm{MSD}}(\alpha) = \left\langle \|\mathfrak{h}_k - \hat{\mathfrak{h}}_{k/k-1}\|^2 \right\rangle \tag{43}$$

where $\langle \|\cdot\| \rangle$ denotes the time average of the Euclidean norm operator. The MSD for each subchannel is accumulated and averaged for 10 000 frames. The overall MSD is then averaged across the RT subchannels.

The SNR is defined per received antenna. Given that σ_n^2 is the AWGN variance at the input of each receiver, then with both the QPSK signals and the overall random components of the multipath rays normalized to unit energy, we have

$$\mathrm{SNR} = 10 \log \left(\frac{(1+K)}{\sigma_n^2} \right). \tag{44}$$

Unless stated otherwise, we assume that the total transmitted power is restricted to unit power and equally allocated between the T antennas.

Fig. 4 shows the MSD behaviors of a VGRLS estimator and a Kalman filter[6] at a normalized fade rate of $f_D T = 0.002$, where f_D is the maximum Doppler frequency. We observe how the polynomial order N, system SNR, and training sequence length L_t affect the MSD. At low SNR, where noise dominates, an estimator with order $N = 0$ has a slightly better MSD than the others since the algorithm then primarily acts as a noise averaging filter [11]. This is also attributed to a smaller norm of the polynomial coefficients for $N = 0$ because a larger norm

[6]We assume $P_a = 3$ for the Kalman filter, i.e., the same as the predictor length P in the simulations.

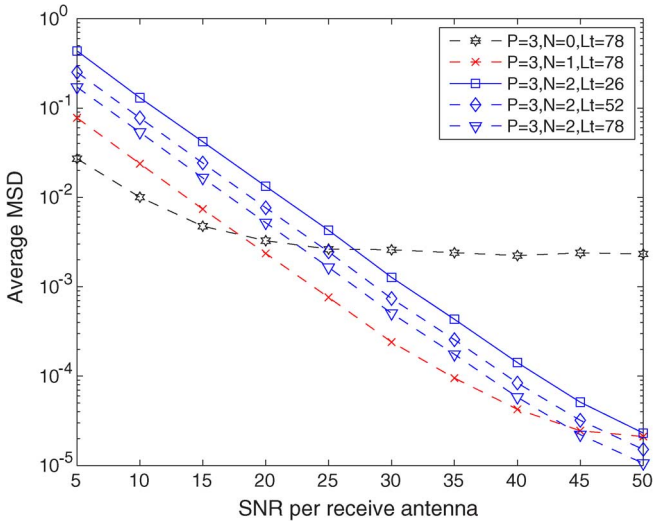


Fig. 5. MSD of the VGRLS estimator for a (2, 2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T$ of 0.0001.

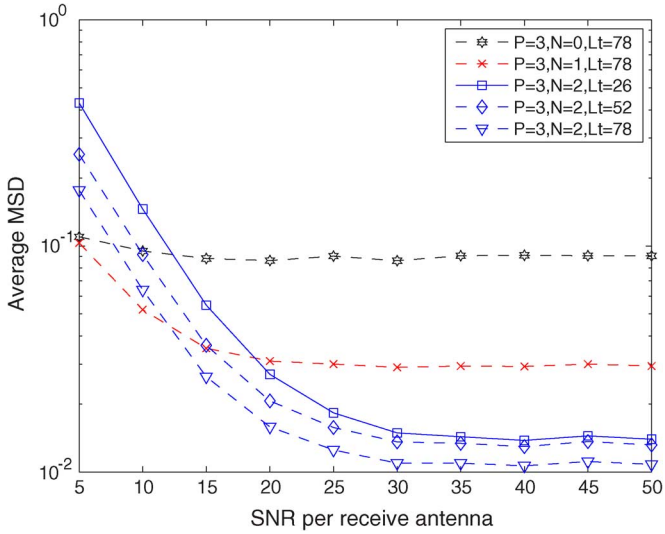


Fig. 6. MSD of the VGRLS estimator for a (2, 2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T$ of 0.01.

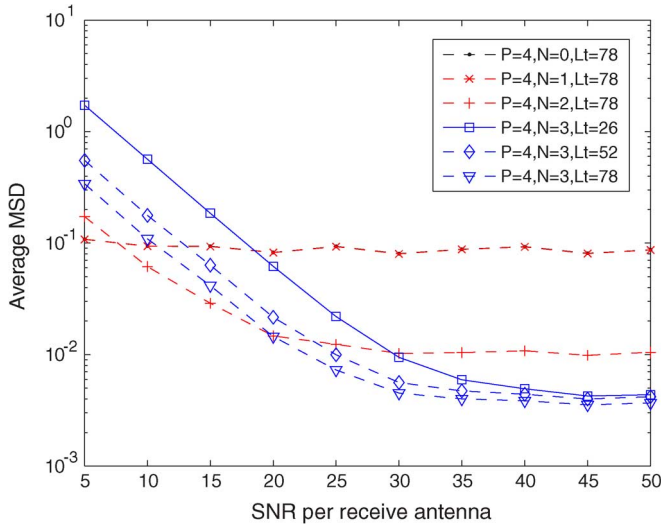


Fig. 7. MSD of the VGRLS estimator for a (2, 2) MIMO system in a Rayleigh fading channel with a normalized fade rate $f_D T$ of 0.01. The VGRLS has a predictor length of $P = 4$ and various order N as shown. (Note that the curves for $N = 0$ and $N = 1$ overlap each other).

"magnifies" the noise. At moderate SNR, an order of $N = 1$ performs not only linear interpolation but also noise averaging and has the lowest MSD. At high SNR, where the effect of modeling error dominates, an order of $N = 2$ has the lowest MSD. These behaviors are consistent with those reported in [11].

It is interesting to note that the estimator MSD exhibits a floor at sufficiently high SNR, regardless of the value of N. This was initially thought to be due to the nonconvergence of the estimator due to an insufficient number of training symbols. However, when the number of training symbols is increased to $L_t = 78$,[7] the floor for $N = 2$ is still visible at very high SNR, although the effect is slight within the observed SNR range. This is unlike the scalar case in [11], where the use of $L_t = 52$ effectively removes the error floor for $N = 2$ at high SNR. As will be shown later, this is actually due to the fading-related error that cannot be reduced by increasing the SNR or the polynomial order N. We also note that increasing the length of the training sequence improves the steady-state MSD performance of the estimator across the SNR region for all the estimator orders.

Fig. 5 shows the MSD of the VGRLS at a slower normalized fade rate of 0.0001, and Fig. 6 shows the MSD at a faster normalized fade rate of 0.01. The results show a similar trend in MSD performance, i.e., at low SNR, there is not much difference between the various orders, but at high SNR, $N = 2$ offers a significantly lower MSD. They show that the VGRLS is able to operate in both slow- and fast-fading environments because it converges in both scenarios.

These figures also reveal the effect of fade rate on the estimator. A faster fading channel is more difficult to track; hence, it introduces a fade-rate-related error. Furthermore, due to the truncation effect of the Taylor series expansion, more terms in the series (hence a higher polynomial order and predictor length) are required to support a higher fade rate [19]. This is shown in Fig. 7, where a VGRLS with $P = 4$ and $N = 3$ produces a lower MSD at a fade rate of 0.01 when compared to Fig. 6. However, we note that in slower fading, a higher polynomial order and predictor length does not offer any advantage. As shown in Fig. 8 for a fade rate of 0.002, a VGRLS with $P = 4$ and $N = 3$ has the same MSD at high SNR as that of $P = 3$ and $N = 2$. We deduce from these results that the "floor" at high SNR is attributed to fade-rate-related error that cannot be lowered by using higher values of P and N. We note that the Kalman filter produces a slightly lower MSD than the VGRLS at these fade rates.

We have assumed a uniform power delay profile in our simulations for simplicity. However, in reality, the power delay profile may not be uniform. We have investigated the MSD performance of VGRLS at a normalized fade rate of 0.002 for a (2, 2) Rayleigh fading with a nonuniform power delay profile modeled according to the SUI-4 channel model [27]. This is a three-ray model with a power profile of 0, −4, and

[7]Data length L_d is still 116 symbols.

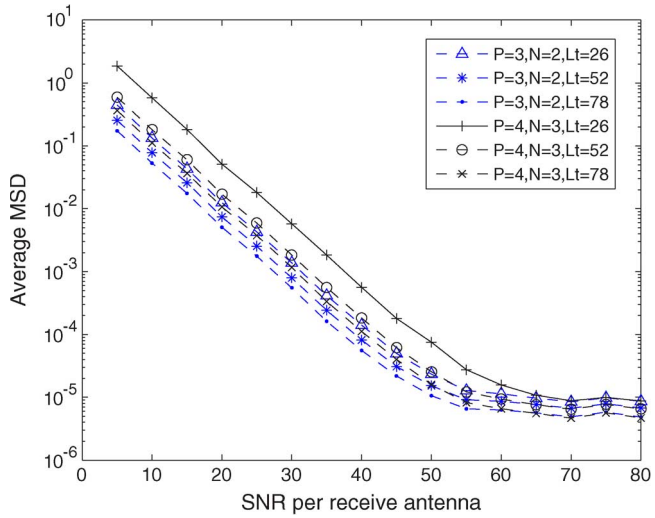


Fig. 8. MSD of the VGRLS estimator for a (2, 2) MIMO system in a Rayleigh fading channel with a normalized fade rate f_DT of 0.002. The VGRLS has a predictor length of $P = 3$, $N = 2$ and $P = 4$, $N = 3$, with various training symbol length.

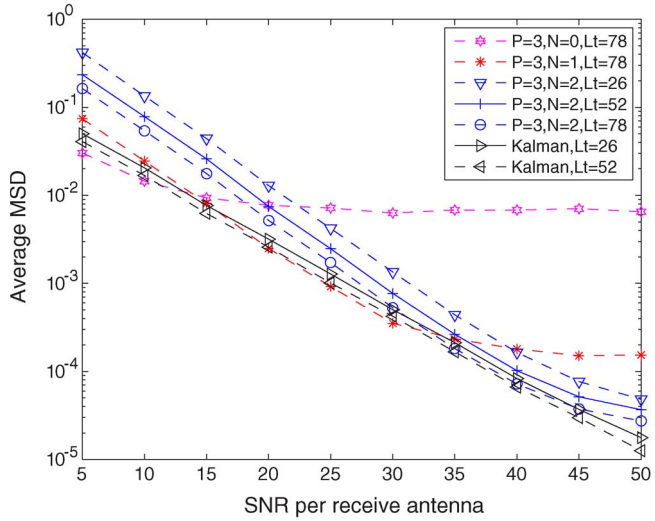


Fig. 9. MSD of the VGRLS estimator for a (2, 2) MIMO system in a Rayleigh fading channel with a normalized fade rate f_DT of 0.002. The VGRLS has a predictor length of $P = 3$ and various N with various training symbol length, as shown. The power delay profile is nonuniform with a power profile of 0, -4, and -8 dB.

-8 dB. The result is shown in Fig. 9. Compared to Fig. 4 for a uniform power delay profile, we note that there is negligible difference in performance. A uniform profile is considered as one of the more severe profiles as all the multipath rays have equal power. It is also used as a test profile for the purpose of GSM's equalizer testing [28]. In the following, a uniform delay profile is used, unless stated otherwise.

Figs. 10 and 11 show the MSD performance of the estimator for a (4, 4) MIMO system at normalized fade rates of 0.002 and 0.0001, respectively. In general, the MSD performance is worse than that of a (2, 2) system, although it is improved with a longer training sequence length L_t. We note the irreducible MSD floor for $N = 2$ with $L_t = 26$. This appears to be due to the failure of the estimator to converge within 26 symbol periods.

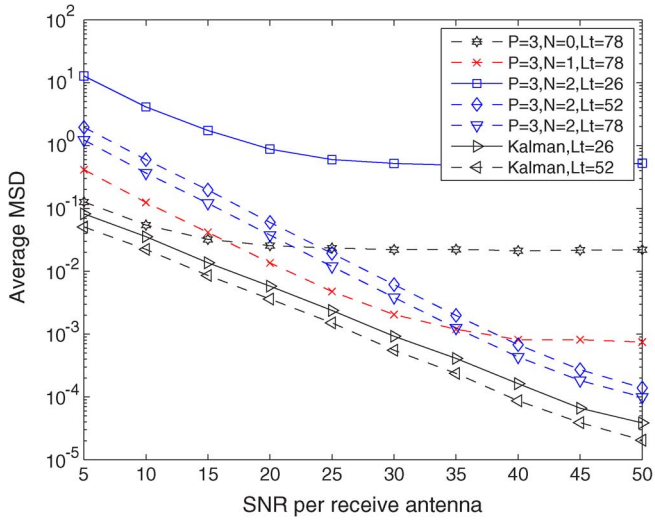


Fig. 10. MSD of the VGRLS estimator and a Kalman filter for a (4, 4) MIMO system in a Rayleigh fading channel with a normalized fade rate f_DT of 0.002.

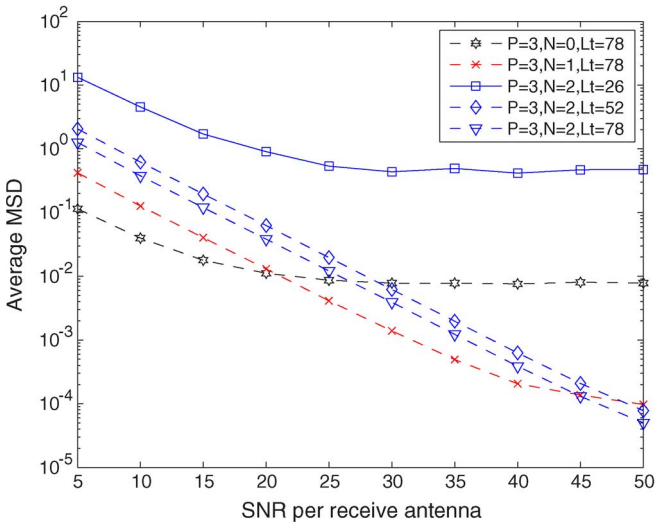


Fig. 11. MSD of the VGRLS estimator for a (4, 4) MIMO system in a Rayleigh fading channel with a normalized fade rate f_DT of 0.0001.

We now consider the error rate performance of an integrated receiver consisting of the VGRLS estimator and a vector DFE [23] operating in a decision-directed mode following training. The estimated channel responses from the VGRLS estimator are used to calculate the tap coefficients of the DFE, and the outputs of the DFE are used by the estimator to update the estimated channel responses. The simulation at each SNR point is carried out until 200 symbol errors are encountered, and the symbol error rate (SER) is averaged across the T transmitted signal streams. The simulation for the Kalman-filter-based receiver follows the same approach.

Figs. 12–14 show the average SER performance of independently transmitted QPSK signal streams in (2, 2) Rayleigh fading, (4, 4) Rayleigh fading, and (2, 2) Rician fading MIMO systems. A MIMO MMSE-DFE with $N_f = 4$ feedforward filter taps, $N_b = 2$ feedback filter taps, and a decision delay of $\triangle = 3$ is used. The taps are symbol spaced, and frames with a training sequence of $L_t = 78$ symbols and a data frame of $L_d = 1160$ symbols are used. The result shows that the receiver is able

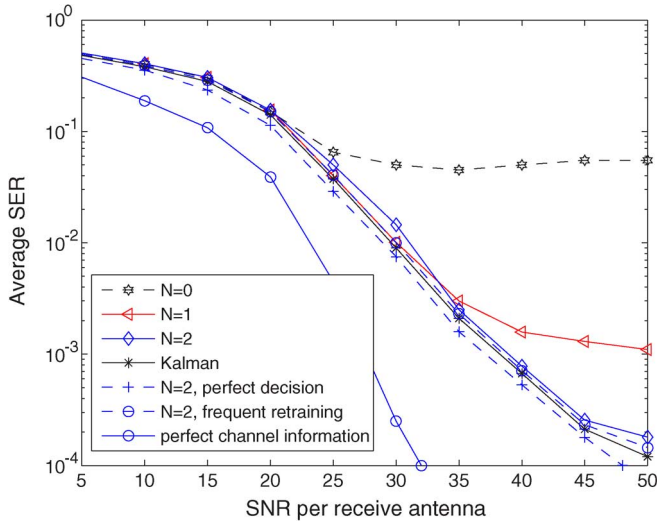


Fig. 12. Average SER performance of the VGRLS estimator and a Kalman filter for a (2, 2) MIMO VBLAST-type system in a Rayleigh fading channel with a normalized fade rate $f_D T$ of 0.002 using a MIMO MMSE DFE.

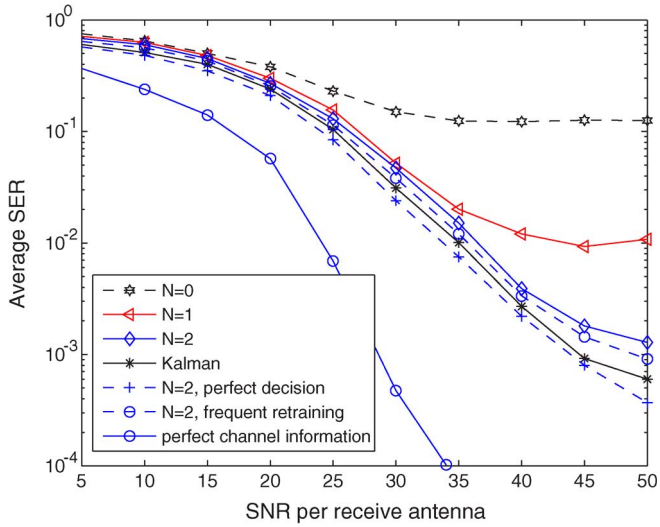


Fig. 13. Average SER performance of the VGRLS estimator and a Kalman filter for a (4, 4) MIMO VBLAST-type system in a Rayleigh fading channel with a normalized fade rate $f_D T$ of 0.002 using a MIMO MMSE DFE.

to track the channel over a reasonably long data frame before the next training phase in the subsequent frame. For $N = 2$, we have also simulated a more frequent periodic retraining using $L_t = 78$ symbols in the first frame and $L_t = 26$ in all the subsequent frames with $L_d = 116$ in all the frames. The results show that a more frequent periodic retraining offers only marginal improvement in error rate performance for the scenarios considered, although the improvement appears to be slightly greater for the (4, 4) case.

The Rayleigh simulations each have a total of unit transmit power equally shared among the transmitters with $K = 0$ and a normalized fade rate of 0.002. From Figs. 12 and 13, we observe that at low to moderate SNR, all the channel estimators lead to comparable SER performance, regardless of the polynomial order used. However, at high SNR, $N = 2$ performs somewhat better and is comparable to that of the Kalman-filter-based receiver, which also starts to exhibit an error floor that

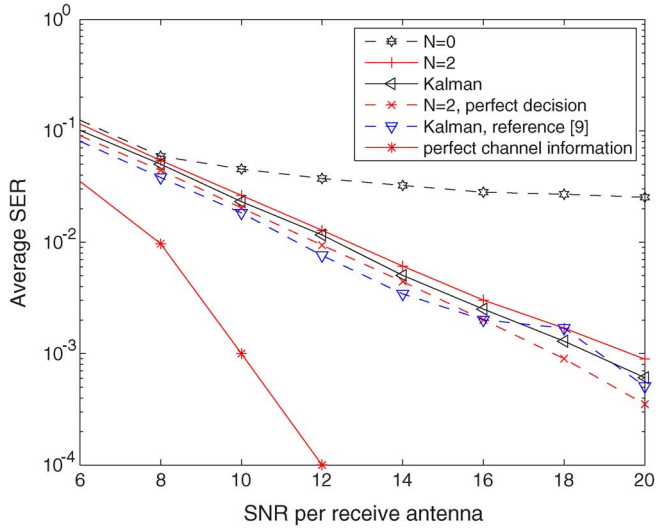


Fig. 14. Average SER performance of the VGRLS estimator and a Kalman filter for a (2, 2) MIMO VBLAST-type system in a Rician fading channel with a normalized fade rate $f_D T$ of 0.007 and $K = 10$. The Kalman filter result in [9] is also plotted for reference.

is not much different from that seen when using the VGRLS-based receiver. We note that at a SER of 10^{-3}, there is a 5-dB difference in the Kalman filter's performance between a (2, 2) and (4, 4) system. Results with perfect decision feedback (i.e., using known transmitted symbols) to the estimator and perfect channel information for the DFE tap calculation when using the VGRLS estimator are also included and show that the $N = 2$ case suffers only modest losses.

We follow the approach in [9] for the simulation of the (2, 2) Rician fading. We allocate a unit transmit power to each of the transmitters, so the resulting graph has a $\log_{10}(T) = 3$ dB increase in the SNR per antenna compared to when the total transmit power is limited to a unit energy. A Rician K-factor of 10 and a normalized fade rate of 0.007 are used. We also assume the specular components of the fading channel responses to be known when simulating the Kalman filter. This simplifies the simulation by not requiring the state transition matrix to be restructured [15]. However, we have used three instead of two multipath rays in each subchannel, and this affects the vector DFE's design. We note that the resulting Kalman filter's curve is reasonably close to that in [9].

For the VGRLS estimator, the specular components of the fading channel responses are not known and are estimated, together with the diffuse components. From Fig. 14, we note that at a SER of 10^{-3}, the VGRLS/DFE combination with $N = 2$ is only 1 dB from the Kalman/DFE receiver. Results using perfect decision feedback to the estimators and perfect channel information for the DFE tap calculations are also included. They indicate almost a 9-dB loss with respect to the perfect channel information case at a SER of 10^{-3} but only very modest losses with respect to a Kalman estimator using perfect decision feedback.

VI. Conclusion

We have developed a symbol-by-symbol-based MIMO receiver design with integrated channel estimation and tracking.

In particular, we have developed and used a MIMO channel estimator employing a VGRLS algorithm that is capable of tracking time-varying frequency-selective channel responses in both Rician and Rayleigh fading environments. The estimator does not require channel and noise statistics to operate and does not require specific modeling of the specular component to operate in a Rician fading channel. It has been shown to be robust and can operate in both fast- and slow-fading environments. An MMSE DFE, whose matrix tap coefficients are derived using the channel estimates from the VGRLS estimator, is used as an equalizer. A simple polynomial-based channel prediction module is used to compensate the time lag due to the decision delay of the equalizer. The resulting symbol error performance in Rician and Rayleigh fading channels is shown to be within 1–3 dB of that obtained using an optimum Kalman-based estimator.

ACKNOWLEDGMENT

Y. H. Kho would like to thank W. S. Leon for his advice and consultation during the early stage of this paper.

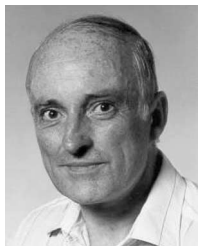
REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
[3] G. J. Foschini and M. J. Gans, "Layered space–time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Lab. Tech. J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.
[4] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.
[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
[6] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[7] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[8] B. D. Hart and D. P. Taylor, "Maximum-likelihood synchronization, equalization, and sequence estimation for unknown time-varying frequency-selective Rician channels," *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 211–221, Feb. 1998.
[9] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
[10] M. Enescu, M. Sirbu, and V. Koivunen, "Adaptive equalization of time-varying MIMO channels," *Signal Process.*, vol. 85, no. 1, pp. 81–93, Jan. 2005.
[11] W. S. Leon and D. P. Taylor, "The polynomial-based generalized recursive least squares estimator for Rayleigh fading channels," in *Proc. GLOBECOM*, 2003, vol. 5, pp. 2401–2405.
[12] C. L. Miller, D. P. Taylor, and P. T. Gough, "Estimation of co-channel signals with linear complexity," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1997–2005, Nov. 2001.
[13] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, vol. COM-11, no. 4, pp. 360–393, Dec. 1963.
[14] *Microwave Mobile Communications*, W. C. Jakes, Ed. New York: Wiley, 1974.
[15] L. M. Davis, I. B. Collings, and R. J. Evans, "Coupled estimators for equalization of fast-fading mobile channels," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1262–1265, Oct. 1998.
[16] G. C. Reinsel, *Elements of Multivariate Time Series Analysis*. New York: Springer-Verlag, 1993.
[17] P. T. Harju and T. I. Laakso, "Polynomial predictors for complex-valued vector signals," *Electron. Lett.*, vol. 31, no. 19, pp. 1650–1652, Sep. 1995.
[18] D. K. Borah and B. D. Hart, "A robust receiver structure for time-varying, frequency-flat, Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 360–364, Mar. 1999.
[19] W. S. Leon and D. P. Taylor, "The generalized polynomial predictor based receiver for the nonselective fading channel," in *Proc. GLOBECOM*, 2000, vol. 2, pp. 927–931.
[20] W. S. Leon, "Equalization and estimation for fading channels," Ph.D. dissertation, Univ. Canterbury, Christchurch, New Zealand, 2003.
[21] A. P. Clark, *Adaptive Detectors for Digital Modems*. London, U.K.: Pentech, 1989.
[22] W. S. Leon and D. P. Taylor, "Steady-state tracking analysis of the RLS algorithm for time-varying channels: A general state-space approach," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 236–238, May 2003.
[23] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.
[24] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA: McGraw-Hill, 2001.
[25] W. van Etten, "Maximum likelihood receiver for multiple channel transmission systems," *IEEE Trans. Commun.*, vol. COM-24, no. 2, pp. 276–283, Feb. 1976.
[26] D. Verdin and T. C. Tozer, "Generating a fading process for the simulation of land-mobile radio communications," *Electron. Lett.*, vol. 29, no. 23, pp. 2011–2012, Nov. 1993.
[27] V. Erceg, *Channel Models for Fixed Wireless Applications*, Feb. 2001. IEEE 802.16 Broadband Wireless Access Working Group.
[28] W. H. Tranter, K. S. Shanmugam, T. S. Rappaport, and K. L. Kosbar, *Principles of Communication Systems Simulation With Wireless Applications*. Upper Saddle River, NJ: Prentice-Hall, 2004.

Yau Hee Kho (S'06) was born in Kuching, Sarawak, Malaysia, in 1974. He received the B.Eng. degree (First-Class Honors) in 1997 from the University of Canterbury, Christchurch, New Zealand, where he is currently working toward the Ph.D. degree.

From 1998 to 2002, he was an Electrical Design Engineer in Singapore. His research interests include estimation and equalization of digital wireless communications, in particular for MIMO systems.

Mr. Kho is a member of the Institution of Engineering and Technology (IET, formerly the IEE, U.K.), where he currently serves as a member of Council and of the Technical and Professional Services Board. He was awarded first prize for the IET Write Around The World (WATW) competition in 2001 and third prize for the postgraduate section of the IEEE Region 10 Student Paper Contest in 2007.



Desmond P. Taylor (M'65–SM'90–F'94–LF'07) was born in Noranda, QC, Canada, on July 5, 1941. He received the B.Sc. (Eng.) and M.Sc. (Eng.) degrees from Queen's University, Kingston, ON, Canada, in 1963 and 1967, respectively, and the Ph.D. degree in electrical engineering from McMaster University, Hamilton, ON, in 1972.

From July 1972 to June 1992, he was with the Communications Research Laboratory and Department of Electrical Engineering, McMaster University. Since July 1992, he has been with the University of Canterbury, Christchurch, New Zealand, where he is the Tait Professor of communications. He is the author or coauthor of approximately 220 published papers and is the holder of two U.S. patents in spread-spectrum communications. His research interests are centered on digital wireless communications systems with primary focus on the development of robust bandwidth-efficient modulation and coding techniques and the development of iterative algorithms for the joint equalization and decoding of fading dispersive channels typical of mobile radio communications. His secondary interests include problems in synchronization, multiple access, and networking.

Prof. Taylor is a Fellow of the Royal Society of New Zealand, the Engineering Institute of Canada, and the Institute of Professional Engineers of New Zealand. One of his papers won the S.O. Rice Award for the best Transactions paper in Communication Theory for 2001.