Signal design and receiver dimensioning for space–time Viterbi equalisation

C.-S.Chou and D.W.Lin

Abstract: Space–time Viterbi equalisation for wireless communication has received much interest recently. The authors consider the associated signal and receiver design for improved transmission performance at high baud rates. Three techniques are proposed; two concern signal design and one concerns receiver design. First, special kinds of sequences called the min-norm training sequences are used for channel estimation. These sequences can minimise the mean-square estimation error in uncorrelated AWGN environments. Secondly, the authors consider using unequal power levels for the training signal and the data signal, with a higher power level for the former. They derive a mathematical expression for the optimal power ratio. Thirdly, a channel estimation method using reduced channel length is proposed. This method can reduce the channel estimation error in low SNR environments. Small-scale Monte Carlo simulations are conducted to investigate the performance gain of these techniques in wireless transmission. The results show varying degrees of advantage under different conditions.

1 Introduction

In the area of wireless communications, there has been increasing interest in space–time signal processing [1, 2]. Space–time processing is able to enhance the received signal subject to multidirectional, multipath propagation better than time-only processing. It is also able to separate signals travelling in different spatial directions better, facilitating reduced interference in a multiple-access system. These properties are of increasing importance in view of the increasing transmission rates of new wireless systems and the expected growth in wireless traffic.

Among the space–time processing schemes, space–time Viterbi equalisation has been subject to much recent research. In this paper, we consider signal and receiver design for space–time Viterbi equalisation in high-speed nonspread-spectrum wireless communication and evaluate the resulting performance using computer simulation. Fig. 1 shows the structure of the transmission system, where the space–time Viterbi equaliser is composed of a vector channel estimator and a vector-channel Viterbi sequence estimator. The system employs a training sequence for the equaliser's use, but the receiver does not have to know the received signal's direction of arrival (DOA) or the array manifold vector.

We propose three techniques to improve the transmission performance under the above receiver architecture. Two concern signal design and the third concerns receiver design. First, we consider the design of the training sequence for channel estimation. A design that minimises the mean-square estimation error is presented. Secondly, we consider using unequal power levels for the training sequence and the data signal to effect a more accurate channel estimate for the benefit of data transmission performance. Thirdly, we consider the use of reduced channel lengths in channel estimation and Viterbi equalisation. We investigate the relationship between tap choice and estimation performance, and we demonstrate that reduced-length channel estimation can be advantageous especially in low SNR environments.

As shown in Fig. 1, let a_i be the ith transmitted (baseband) symbol, $p(t)$ be the pulse shape, $q(t)$ be the impulse response of the receiver's front-end filters, M be the number of elements in the receiver's antenna array, and $\underline{b}(t)$ be the vector channel impulse response. We have

$$\underline{z}(t) = \sum_{i=-\infty}^{\infty} a_i[p(t-iT) * \underline{b}(t) * q(t)] + \underline{n}(t)$$
$$\triangleq \sum_{i=-\infty}^{\infty} a_i \underline{r}(t-iT) + \underline{n}(t) \quad (1)$$

where T is the symbol period and $\underline{r}(t)$ is the combined impulse response of the pulse shaping filter, the vector channel, and the receiver's front-end filters. Sampling of $\underline{z}(t)$ yields

$$\underline{x}(k) = \boldsymbol{R}\underline{a}(k) + \underline{n}(k) \quad (2)$$

where

$$\boldsymbol{R} = [\ldots, \underline{r}(kT), \boldsymbol{r}([k+1]T), \underline{r}([k+2]T), \ldots] \quad (3)$$

i.e. the overall channel impulse response matrix, and

$$\underline{a}(k) = [\ldots, a_k, a_{k-1}, a_{k-2}, \ldots]' \quad (4)$$

where $'$ denotes the matrix transpose. For convenience, we let each row of $\boldsymbol{R}$ have unit energy; that is, the sum of square values of each row of $\boldsymbol{R}$ is equal to one.

Consider tentatively the transmission of the training sequence alone. Let the training sequence be of length L and start at time 1. Let q be the length of the sampled vector impulse response $\underline{r}(kT)$, that is, q is the number of columns in $\boldsymbol{R}$. In addition, assume for simplicity that $q \leq L$.

IEE Proceedings online no. 20010271

DOI: 10.1049/ip-com:20010271

Paper first received 20th December 1999 and in revised form 30th November 2000

The authors are with the Department of Electronics Engineering and Center for Telecommunications Research, National Chiao Tung University, Hsinchu, Taiwan 30010, Republic of China

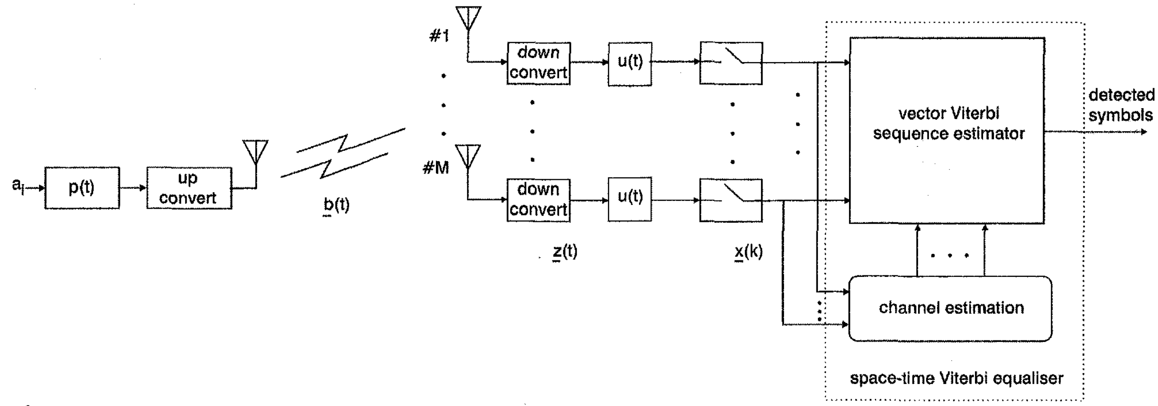


Fig. 1 *Wireless transmission system with space–time Viterbi equalisation*

Then the centre section of the received data, where there is full convolution between the overall channel impulse response and the training sequence, has length $L - q + 1$ and starts at time q. And we may express it in matrix form as [1]

$$\boldsymbol{X} = \boldsymbol{RG} + \boldsymbol{N} \tag{5}$$

where

$$\boldsymbol{X} = [\underline{\boldsymbol{x}}(q), \underline{\boldsymbol{x}}(q+1), \ldots, \underline{\boldsymbol{x}}(L)] \tag{6}$$

$\boldsymbol{N}$ is a matrix of noise samples, and $\boldsymbol{G}$ is the training symbol matrix given by

$$\boldsymbol{G} = [\underline{\boldsymbol{g}}_1', \underline{\boldsymbol{g}}_2', \ldots, \underline{\boldsymbol{g}}_q']' \tag{7}$$

with

$$\underline{\boldsymbol{g}}_i = [a_{q-i+1}, q_{q-i+2}, \ldots, a_{L-i+1}] \tag{8}$$

We assume the noise into the different antennas to be zero-mean white, uncorrelated, and with an identical variance σ^2.

2 Min-norm training sequences

Let $\hat{\boldsymbol{R}}$ denote the channel response estimate. Consider the least-squares estimation of the channel response from the training symbols as

$$\min_{\hat{\boldsymbol{R}}} \|\boldsymbol{X} - \hat{\boldsymbol{R}}\boldsymbol{G}\|_F^2 \tag{9}$$

where $\|\cdot\|_F$ denotes the Frobenius norm [3]. Setting the derivatives with respect to the unknowns to zero, we obtain

$$\hat{\boldsymbol{R}} = \boldsymbol{X}\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1} \tag{10}$$

where the superscript H denotes the Hermitian transpose. Thus the sum-squares error in channel estimation is given by

$$\begin{aligned} \|\hat{\boldsymbol{R}} - \boldsymbol{R}\|_F^2 &= \|(\boldsymbol{RG} + \boldsymbol{N})\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1} - \boldsymbol{R}\|_F^2 \\ &= \|\boldsymbol{N}\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}\|_F^2 \end{aligned} \tag{11}$$

It is shown in the Appendix that minimisation of the mean-square estimation error is equivalent to minimisation of the following quantity:

$$\sum_{i=1}^{q} \frac{1}{|\sigma_i|^2} \tag{12}$$

where σ_i are the singular values of $\boldsymbol{G}$. For convenience, term the quantity in eqn. 12 the normalised error norm and term sequences minimising it the min-norm sequences.

Note that, for any sequence, the normalised error norm can be reduced by simply increasing the sequence's energy. This trivial implication is certainly not what the above result is useful for. Therefore, in actual use of the result, we should minimise the normalised error norm subject to an energy constraint on the sequence.

An interesting question is whether some of the well known kinds of sequences in communications, such as the maximum-length (ML) sequences [4], would have the min-norm property or be nearly so. While a theoretical derivation is not conducted in this work, the simulation results presented later show that ML sequences can yield significantly higher estimation error than a sequence which minimises the normalised error norm. ML sequences are also less flexible in their lengths, which can only be equal to $2^n - 1$ where n is an integer, whereas a min-norm sequence can be of any length.

For want of more precise characterisations of the min-norm sequences, they need to be found by an exhaustive search over all sequences of the considered length for the considered modulation method. A nonexhaustive search may be substituted to obtain suboptimal training sequences when the number of possible training sequences makes exhaustive search infeasible. Note that the search is conducted offline at the designing stage of the transmission system. It does not affect the computational complexity of the transceiver in actual signal transmission.

The channel estimation computation (eqn. 10) requires matrix inversion, in principle. If needed to be carried out in real-time, its computational complexity could be a concern in practical implementation. However, since the matrix that needs be inverted is solely a function of the training sequence and hence is known in advance, its inverse can be precomputed and stored. Indeed, the whole multiplicative factor $\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}$ can be computed in advance. Thus matrix inversions do not constitute a part of the per-sample computational complexity, but only some matrix multiplications. The reduced-length channel estimation discussed later is similar.

An independent study on training sequence design was reported in [5]. There the authors concentrated on single-antenna reception and (primarily) BPSK modulation. Our study also considers QPSK. Numerical results show that, for transmission systems employing QPSK modulation, BPSK-based min-norm sequences may yield more noisy channel estimates than QPSK-based ones.

3 Unequal power levels for training and data signals

Wireless communication systems are often designed to transmit a predefined sequence periodically for purposes of synchronisation and/or equaliser training. In usual system design, the percentage of time occupied by the training signal is much smaller than that by the data signal. Since the training signal determines the accuracy in channel estimation and thereby the subsequent equalisation performance for the data signal, it appears that by increasing the relative power level of the training signal in comparison to that of the data signal (with the total transmission power kept the same), we may effect a better channel estimation performance and consequently a better data transmission performance.

To investigate the possible gain with this arrangement, let the number of data symbols be D and let X_d denote the received data signal matrix. Then we have

$$\boldsymbol{X}_d = \boldsymbol{R}\boldsymbol{S} + \boldsymbol{N}_d \tag{13}$$

where $\boldsymbol{S}$ is the matrix of transmitted data and $\boldsymbol{N}_d$ is the associated matrix of additive noise samples. $\boldsymbol{S}$ is of Toeplitz form as

$$\boldsymbol{S} = \begin{bmatrix} s_{q-1} & s_q & s_{q+1} & \cdots \\ s_{q-2} & s_{q-1} & s_q & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ s_0 & s_1 & s_2 & \cdots \end{bmatrix} \tag{14}$$

The Viterbi algorithm for maximum-likelihood sequence estimation (MLSE) [6] examines, in principle, each possible transmitted data sequence and computes its associated channel output in the absence of noise using the estimated channel response, and selects the sequence whose associated channel output is closest to the actually received data signal. In uncorrelated AWGN channel transmission, therefore, it is desirable to minimise the mean-square error between $\boldsymbol{X}_d$ and $\hat{\boldsymbol{R}}\boldsymbol{S}$, the latter being the estimated received signal in the absence of noise with $\boldsymbol{S}$ as the channel input. Now since

$$\boldsymbol{X}_d = \hat{\boldsymbol{R}}\boldsymbol{S} + \boldsymbol{R}\boldsymbol{S} + \boldsymbol{N}_d - \hat{\boldsymbol{R}}\boldsymbol{S} = \boldsymbol{N}_d - \boldsymbol{N}\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}\boldsymbol{S} \tag{15}$$

we have

$$\begin{aligned} \|\boldsymbol{X}_d - \hat{\boldsymbol{R}}\boldsymbol{S}\|_F^2 &= tr\Big\{[\boldsymbol{N}_d - \boldsymbol{N}\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}\boldsymbol{S}] \\ &\qquad \times [\boldsymbol{N}_d - \boldsymbol{N}\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}\boldsymbol{S}]^H\Big\} \\ &= tr\Big\{\boldsymbol{N}_d\boldsymbol{N}_d^H - \boldsymbol{N}_d\boldsymbol{S}^H[(\boldsymbol{G}\boldsymbol{G}^H)^{-1}]^H\boldsymbol{G}\boldsymbol{N}^H \\ &\qquad - \boldsymbol{N}\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}\boldsymbol{S}\boldsymbol{N}_d^H \\ &\qquad + \boldsymbol{N}\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}\boldsymbol{S}\boldsymbol{S}^H[(\boldsymbol{G}\boldsymbol{G}^H)^{-1}]^H\boldsymbol{G}\boldsymbol{N}^H\Big\} \end{aligned} \tag{16}$$

Assume that the transmitted data are zero-mean, i.i.d., and with variance σ_d^2. Then

$$\begin{aligned} E\{\|\boldsymbol{X}_d - \hat{\boldsymbol{R}}\boldsymbol{S}\|_F^2 &= DM\sigma^2 \\ &\quad + DM\sigma^2\sigma_d^2 \cdot tr\{\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}[(\boldsymbol{G}\boldsymbol{G}^H)^{-1}]^H\boldsymbol{G}\} \\ &= DM\sigma^2[1 + \sigma_d^2 \cdot tr\{(\boldsymbol{G}\boldsymbol{G}^H)^{-1}\}] \end{aligned} \tag{17}$$

where M is the number of array elements and σ^2 the variance of noise samples.

Let σ_t^2 denote the mean-square value of the training symbols. For convenience, define

$$\boldsymbol{G}_N \triangleq \frac{1}{\sigma_t}\boldsymbol{G} \tag{18}$$

that is, $\boldsymbol{G}_N$ is the training data matrix formed of the normalised training sequence. Then

$$E\{\|\boldsymbol{X}_d - \hat{\boldsymbol{R}}\boldsymbol{S}\|_F^2\} = DM\sigma^2\left[1 + \frac{\sigma_d^2}{\sigma_t^2} \cdot tr\{(\boldsymbol{G}_N\boldsymbol{G}_N^H)^{-1}\}\right] \tag{19}$$

Therefore, if we were only to minimise the the mean-square error, then we should let σ_d^2/σ_t^2 approach zero, or equivalently, let all the power be put into transmitting the training data. However, the error rate of Viterbi equalisation is not minimised by minimising the noise, but by minimising the noise at a given signal level, or equivalently by maximising the SNR given by $\sigma_d^2/E\{\|\boldsymbol{X}_d - \hat{\boldsymbol{R}}\boldsymbol{S}\|_F^2\}$.

To proceed, let the total transmission power over a frame of data consisting of L training symbols and D data symbols be K, i.e. $L\sigma_t^2 + D\sigma_d^2 = K$. Further, let g be the ratio of the power of the training signal to that of the data signal, i.e.

$$g \triangleq \sigma_t^2/\sigma_d^2 \tag{20}$$

Then

$$\sigma_d^2 = \frac{K}{Lg + D} \tag{21}$$

and consequently

$$\begin{aligned} &\frac{\sigma_d^2}{E\{\|\boldsymbol{X}_d - \hat{\boldsymbol{R}}\boldsymbol{S}\|_F^2\}} \\ &\quad = \frac{K}{(Lg + D)DM\sigma^2\left[1 + \frac{1}{g} \cdot tr\{(\boldsymbol{G}_N\boldsymbol{G}_N^H)^{-1}\}\right]} \end{aligned} \tag{22}$$

Straightforward calculus leads to the optimal g as

$$g = \sqrt{\frac{D}{L} \cdot tr\{(\boldsymbol{G}_N\boldsymbol{G}_N^H)^{-1}\}} \tag{23}$$

In most systems, D/L is approximately between 5 and 15. As an example, a (nonexhaustively searched) min-norm training sequence for QPSK modulation that we obtained has $tr\{(\boldsymbol{G}_N\boldsymbol{G}_N^H)^{-1}\} \approx 0.3383$. If $D/L = 10$, then the optimal g is about $\sqrt{(10 \times 0.3383)} \approx 1.8393$. In other words, the best performance is obtained when the training signal power is 2.65dB larger than the data signal power.

4 Reduced-length channel estimation

The complexity of Viterbi equalisation grows exponentially with the length of the model channel response used in the equaliser. Therefore, it is desirable to keep the length of this response short. One way to achieve it is to shape the response for a designed length through proper filtering [7, 8]. Some propose considering only the stronger taps in the Viterbi MLSE, where the selected taps need not be contiguous in time [9]. But the discontinuity in tap locations makes the control mechanism in the Viterbi algorithm somewhat complicated. In this work, we let the model channel response in the Viterbi equaliser comprise a contiguous segment of taps, and we investigate the implication of tap choice on receiver performance.

Employing the channel models of [10], we find that most of the discrete mobile channels at a 10^6-baud symbol rate will have over 85% of the channel response energy contained within four taps. In fact, we find that, in low SNR environments, estimation of channel responses using smaller channel lengths may result in a smaller estimation error than using larger channel lengths. This is shown in the following.

To begin, let d be the length of the reduced-length channel response estimate. We shall consider a delayed decision-feedback sequence estimation architecture [12]. Let J be the length of channel response used in the Viterbi equaliser part of the sequence estimator. Thus $J \le d$. We seek to position these J taps where the channel response contains maximum energy. More precisely, let $\hat{\boldsymbol{R}}$ be the least-squares estimate of the full-length channel response as before. Then choose d_1 according to

$$\max_{d_1} \left\{ \sum_{i=d_1}^{d_1+J-1} \underline{\hat{r}}_i^H \underline{\hat{r}}_i \right\} \tag{24}$$

where $\underline{\hat{r}}_i$ is the ith column of $\hat{\boldsymbol{R}}$. Form $\tilde{\boldsymbol{G}}$ out of the d_1th to the $(d_1 + d - 1)$th row of $\boldsymbol{G}$. Then compute $\tilde{\boldsymbol{R}}$ as

$$\tilde{\boldsymbol{R}} = \boldsymbol{X}\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} \tag{25}$$

which is the estimated channel with reduced channel length.

To see the amount of estimation error in $\tilde{\boldsymbol{R}}$, define $d_2 \triangleq q - (d_1 + d - 1)$ for convenience, where q is the number of columns in $\boldsymbol{R}$. Since $\boldsymbol{X} = \boldsymbol{RG} + \boldsymbol{N}$

$$\begin{aligned}\tilde{\boldsymbol{R}} &= (\boldsymbol{RG} + \boldsymbol{N})\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} \\ &= \boldsymbol{RG}\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} + \boldsymbol{N}\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} \\ &= \boldsymbol{R}\begin{bmatrix} 0_{d_1-1} \\ \tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H \\ 0_{d_2} \end{bmatrix}(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} \\ &\quad + \boldsymbol{R}\begin{bmatrix} \underline{g}_1 \\ \vdots \\ \underline{g}_{d_1-1} \\ 0_d \\ \underline{g}_{q-d_2+1} \\ \vdots \\ \underline{g}_q \end{bmatrix}\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} + \boldsymbol{N}\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1}\end{aligned} \tag{26}$$

where 0_k denotes a k-row null matrix of appropriate number of columns and $\underline{g}_i$ is the ith row of $\boldsymbol{G}$. Therefore

$$\tilde{\boldsymbol{R}} = \boldsymbol{R}\begin{bmatrix} 0_{d_1-1} \\ \boldsymbol{I}_{d\times d} \\ 0_{d_2} \end{bmatrix} + error \tag{27}$$

where

$$error = \boldsymbol{R}\begin{bmatrix} \underline{g}_1 \\ \vdots \\ \underline{g}_{d_1-1} \\ 0_d \\ \underline{g}_{q-d_2+1} \\ \vdots \\ \underline{g}_q \end{bmatrix}\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} + \boldsymbol{N}\tilde{\boldsymbol{G}}^H(\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^H)^{-1} \tag{28}$$

We name the first RHS term the model error and the second the noise error. The model error comes from the truncation of channel response length, and the noise error is from the additive noise. When the energy of the truncated channel taps is much smaller than the noise power, the noise error constitutes the dominant error term.

By a similar derivation to that in Section 3, we can show that the Frobenius norm of the noise error is equal to

$$\sum_{i=1}^{d} \frac{1}{|\tilde{\sigma}_i|^2} \tag{29}$$

where $\tilde{\sigma}_i$ are the singular values of $\tilde{\boldsymbol{G}}$. Let σ_i and $\tilde{\sigma}_i$ be arranged in descending numerical order, that is

$$|\sigma_1| > |\sigma_2| > \ldots > |\sigma_q| \tag{30}$$

$$|\tilde{\sigma}_1| > |\tilde{\sigma}_2| > \ldots > |\tilde{\sigma}_d| \tag{31}$$

Then by the interlacing theorem [11], we have

$$|\sigma_i| > |\tilde{\sigma}_i| > |\sigma_{i+q-d}|, \quad i = 1, \ldots, d \tag{32}$$

thus

$$\sum_{i=1}^{d} \frac{1}{|\tilde{\sigma}_i|^2} < \sum_{i=1}^{q} \frac{1}{|\sigma_i|^2} \tag{33}$$

From the above discussion, when the SNR is low (larger noise power), using a reduced number of taps for channel estimation can be better. On the other hand, in high SNR environments (noise power is small compared to error in channel truncation), the noise error will not dominate the total estimation error and reduced-length channel estimation may not be better. Simulation shows that in the SNR range of 1–9dB, reduced-length channel estimation often performs better than estimation with nonreduced length, especially when the normalised error norm associated with the training sequence matrix is significantly greater than the minimum.

5 Simulation results

We can simulate space–time Viterbi equalisation. More exactly, we employ the delayed decision-feedback sequence estimation architecture [12]. The receiver's antenna array contains three elements spaced one carrier wavelength apart. The length of channel response estimate used in the MLSE part of the receiver is four taps and that for decision-feedback equalisation (DFE) part is one tap, whether the channel is estimated with or without reduced length. A reason for using a small 4-tap channel length in the MLSE is that, as mentioned in Section 4, this length is enough to capture the majority of the transmitted signal power for most of the channels we are considering. In addition, the number of states in the Viterbi equaliser would be large if a large channel length is used. The branch metric for the Viterbi equaliser used in our simulation is the squared error between the received signal samples and the convolution output of the estimated channel response with the data sequence corresponding to this branch.

In general, if the tap number of the channel employed in the MLSE is J and the modulation is P-ary PSK, then the number of states is P^{J-1}. The number of branch metric computations is P^{J-2} per state per sample. We consider QPSK modulation, hence $P = 4$. Since we have $J = 4$, the number of states is 64 and the number of branch metric computations is 16 per state per sample. And the total number of branch metric computations per sample is thus 1024.

To evaluate the performance of the proposed techniques on the above space–time Viterbi equaliser, 100 channel responses are generated according to the model of [10]. The channels are time-invariant for simplicity in simulation, although the theory presented so far in this paper does not make that assumption. The average performance of different techniques over this set of channels is obtained. The model of [10] covers different propagation environments, including urban, suburban, and hilly. The channels generated have RMS delay spreads between about 1 and 10 microseconds. The QPSK signals ($a_i \in \{\pm 1, \pm j\}$) are transmitted at 1 Mbaud with raised-cosine pulse shaping with $\alpha = 0.75$. We assume perfect carrier and timing recovery. To facilitate computation, in our simulation the channel impulse response matrix $\boldsymbol{R}$ is truncated to contain 99.9% of the power in the overall channel response $\underline{r}(kT)$ (see Section 1).

First, consider the condition where the training signal and data signal have equal power. We examine the performance of different training sequences and the effects of the length of channel response estimate. Consider using a 15×15 training symbol matrix $\boldsymbol{G}$. To fill it takes 29 training symbols. Disregarding the sequences which are different by a constant scaling factor of -1 or $\pm j$, we have 4^{28} possible QPSK training sequences, which are too many to be searched exhaustively for the min-norm one. By nonexhaustive search, we find that the following sequence has a low normalised error norm (equal to 1.058):

$$1, -1, -j, -1, -1, -1, -j, 1, 1, -j, -1, -1,$$
$$1, -1, j, 1, j, -1, j, j, j, -1, -j, -1, -1, j, j, -j, j$$

For ease of reference, this sequence will be referred to as a quasi-min-norm sequence. In comparison, the normalised error norm associated with the min-norm BPSK sequence (obtained by exhaustive search) is 1.152.

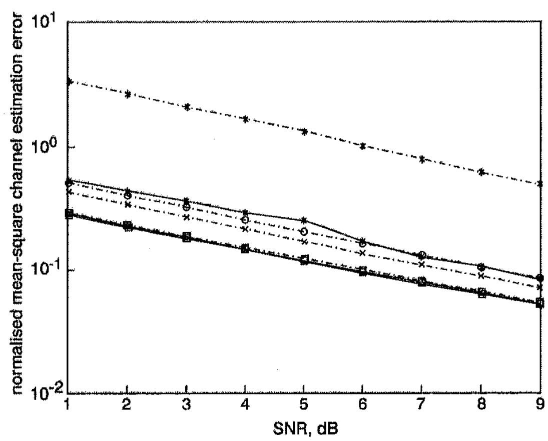


Fig. 2 *Normalised mean-square channel estimation error (3-antenna average) from using different training sequences with or without reduced-length estimation*
3 antennas, 100-channel average, 10^4 symbols per channel run
-·O-· ML15 training at non-reduced length
—O— ML15 training at reduced length
-·✳-· norm = 10.812 training at non-reduced length
—✳— norm = 10.812 training at reduced length
-·×-· norm = 1.614 training at non-reduced length
—×— norm = 1.614 training at reduced length
-·□-· norm = 1.058 training at non-reduced length
—□— norm = 1.058 training at reduced length

Fig. 2 shows some numerical results on the mean-square channel estimation error, normalised with respect to channel transmission gain, using different training sequences with reduced or nonreduced channel lengths. Besides the quasi-min-norm training sequence, we also considered a ML sequence of period 15 (the generating polynomial being $D^4 + D + 1$ and the binary symbols mapped to ± 1) as well as two arbitrarily chosen non-min-norm training sequences. In using the ML sequence it was duplicated once to make it long enough to fill the training symbol matrix. The ML sequence-based training sequence has the advantage of low computational complexity, because the entries in $\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{G}^H)^{-1}$ are either 0 or 1/8. The normalised correlation between different rows of $\boldsymbol{G}$ are very small, being equal to 1/15. The associated normalised error norm is equal to 1.875. The normalised error norms associated with the two arbitrary sequence matrices are 1.614 and 10.812, respectively. The results in Fig. 2 confirm the superiority of the min-norm training sequence over the training sequences of greater normalised error norms in non-reduced-length channel estimation. They also confirm the superiority of reduced-length channel estimation over non-reduced-length estimation in a low SNR environment, especially when the training sequences are associated with greater normalised error norms. The quasi-min-norm sequence performs closely in non-reduced-length and reduced-length estimation, and the performance is close to that of reduced-length estimation of the ML sequence and the sequence with norm 1.614.

Some results on the average symbol error rates (SERs) are shown in Fig. 3. We see (not too surprisingly) increasingly significant improvement by using the quasi-min-norm training sequence compared to training sequences with increasingly greater normalised error norms. In addition, in the case of a non-min-norm sequence (such as the sequence with norm 1.614), the SER performance with reduced-length channel estimation can be significantly better than non-reduced-length estimation (up to about 0.7 dB in this case).

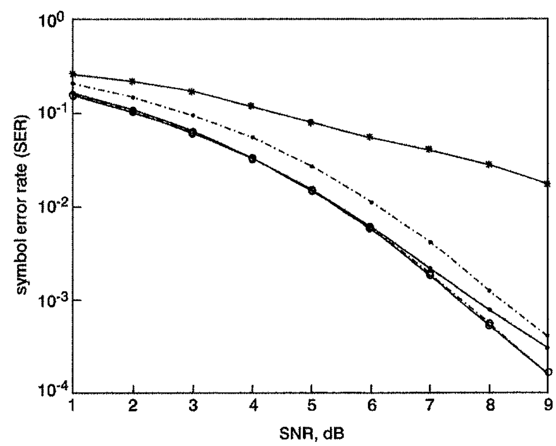


Fig. 3 *Performance of different training sequences with or without reduced-length channel estimation*
3 antennas, 100-channel average, 10^4 symbols per channel run
—✳— sequence norm = 10.812 at reduced length
-·●-· sequence norm = 1.614 at non-reduced length
—●— sequence norm = 1.614 at reduced length
-·O-· sequence norm = 1.058 at non-reduced length
—O— sequence norm = 1.058 at reduced length

Next consider the effect of different power levels for the training signal and the data signal. We simulate several cases of power ratios between 0 and 5 dB. Fig. 4 shows the resulting average SER values at SNR = 5 dB, employing the quasi-min-norm training sequence and the sequence with norm 1.614. We see that the numerical data for the quasi-min-norm sequence verify our earlier theoretical result in Section 3 for the simulated conditions, that a power ratio of 2.65 dB is nearly optimal. For the sequence with norm 1.614, the theoretically optimal power ratio is $g = \sqrt{(10 \times 0.35)} \approx 1.8708$, or 2.72 dB. The numerical data are again corroborative. Figs. 5*a* and *b* further compare the resulting average SERs using equal and unequal training and data signal power levels, at different SNR values and with reduced-length channel estimation, for the quasi-min-norm training sequence and for the training sequence with norm 1.614. Also shown for comparison are results from non-reduced-length channel estimation with equal training and data signal power. The use of unequal power levels yielded consistently better performance.

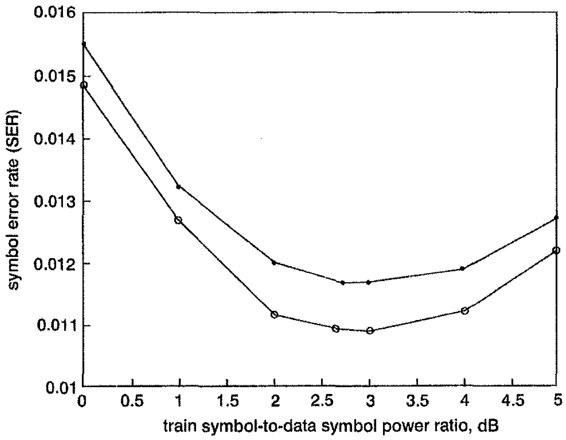
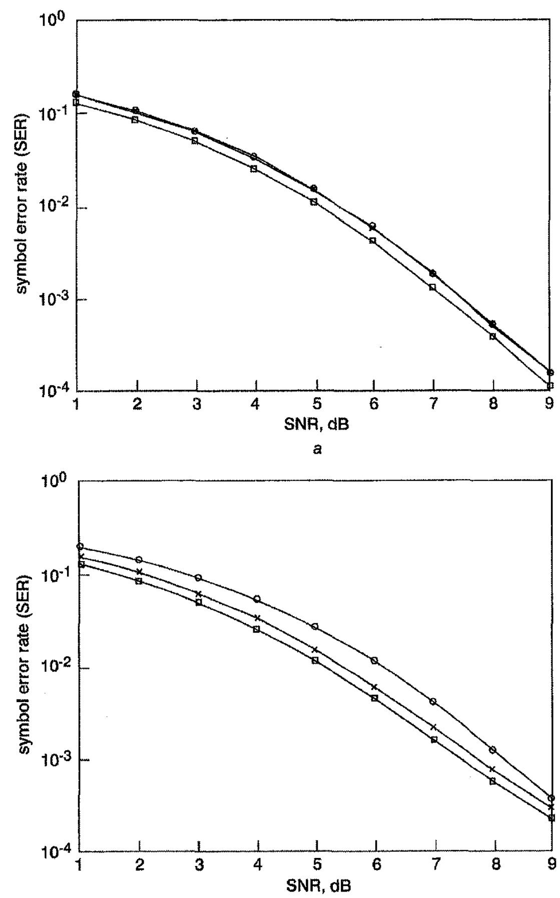


Fig. 4 *Transmission performance at different power ratios between the training signal and the data signal*
3 antennas, 100-channel average, 10^4 symbols per channel run
—●— sequence norm = 1.614, reduced length, $Lv = 4$, $Ldfe = 1$, SNR = 5dB
—○— sequence norm = 1.058, reduced length, $Lv = 4$, $Ldfe = 1$, SNR = 5dB



b

Fig. 5 *Transmission performance of different schemes at different SNR values*
3 antennas, 100-channel average, 10^4 symbols per channel run
a With quasi-min-norm training sequence (normalised error norm = 1.058)
—○— non-reduced length, train/data symbol power ratio = 0dB
—×— reduced length, train/data symbol power ratio = 0dB
—□— reduced length, train/data symbol power ratio = 2.65dB
b With training sequence of normalised error norm 1.614
—○— non-reduced-length, train/data symbol power ratio = 0dB
—×— reduced-length, train/data symbol power ratio = 0dB
—□— reduced-length, train/data symbol power ratio = 2.72dB

We have not investigated in great detail the performance of the proposed techniques in co-channel interference (CCI). Preliminary findings indicate that, for single-user detection with CCI treated as additive noise, the effect of CCI is similar to that of uncorrelated AWGN. More extensive results await further work.

6 Conclusions

We have considered signal and receiver design for space–time Viterbi equalisation in nonspread-spectrum wireless communication at high baud rates, where the Viterbi equaliser is of the delayed decision-feedback sequence estimation variant. The examined system structure employed a training sequence for channel estimation and used the estimate in space–time Viterbi equalisation. Three techniques were proposed to improve the transmission performance, of which two concerned signal design and the third concerned receiver design. On signal design, we first derived a condition for the training sequence which minimised the channel estimation error, and obtained a quasi-optimal sequence for QPSK modulation. Simulation results verified the superiority of this sequence in channel estimation and transmission performance. Secondly, we considered using unequal power levels for the training signal and the data signal. We derived a mathematical expression for computing the best power ratio. Simulation results confirmed that such unequal power arrangement could enhance the transmission performance and that the theoretically derived best power ratio indeed yielded nearly the best performance. For receiver design, we considered using reduced channel lengths to perform channel estimation. We showed that if the resulting modelling error was relatively small compared to noise-induced error, then this could enhance the Viterbi equaliser's performance. Simulation results also confirmed this point.

7 Acknowledgment

This work was supported by the National Science Council of the Republic of China under grant NSC 88-2213-E-009-096.

8 References

1 PAULRAJ, A.J., and PAPADIAS, C.B.: 'Space–time processing for wireless communications', *IEEE Signal Process. Mag.*, 1997, **14**, (6), pp. 49–83
2 MONZINGO, R.A., and MILLER, T.W.: 'Introduction to adaptive arrays' (New York: Wiley, 1980)
3 LANCASTER, P., and TISMENETSKY, M.: 'The theory of matrices with application' (Academic Press, 1985, 2nd edn.)
4 SARWATE, D.V., and PURSLEY, M.B.: 'Crosscorrelation properties of pseudorandom and related sequences', *Proc. IEEE*, 1980, **68**, (5), pp. 593–519
5 CROZIER, S.N., FALCONER, D.D., and MAHMOUD, S.A.: 'Least sum of squared errors (LSSE) channel estimation', *IEE Proc. F*, 1991, **138**, (4), pp. 371–378
6 FORNEY, G.D.: 'Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference', *IEEE Trans. Inf. Theory*, 1972, **IT–18**, pp. 363–378
7 FALCONER, D.D., and MAGEE, F.R.: 'Adaptive channel memory truncation for maximum likelihood sequence estimation', *Bell Syst. Tech. J.*, 1973, **52**, (9), pp. 1541–1561
8 AI-DHAHIR, N., and CIOFFI, J.M.: 'Efficiently computed reduced-parameter input-aided MMSE equalizers for ML detection: a unified approach', *IEEE Trans. Inf. Theory*, 1996, **42**, (3), pp. 903–915
9 ISHII, N., and KOHNO, R.: 'Tap selectable Viterbi equalization in spatial and temporal domains for multipath channel'. Proceedings of IEEE international conference on *Universal personal communications*, 1995, pp. 904–908
10 HASHEMI, H.: 'Simulation of the urban radio propagation channel', *IEEE Trans. Veh. Technol.*, 1979, **VT–28**, (3), pp. 213–225
11 HORN, R.A., and JOHNSON, C.R.: 'Matrix analysis' (Cambridge University Press, 1980)
12 DUEL-HALLEN, A., and HEEGARD, C.: 'Delayed decision-feedback sequence estimation', *IEEE Trans. Commun.*, 1989, **37**, (5), pp. 428–436

9 Appendix

In this Appendix, we show that minimisation of the mean-square estimation error $\|NG^H(GG^H)^{-1}\|_F^2$ is equivalent to minimisation of the normalised error norm $\Sigma_{i=1}^{q} 1/|\sigma_i|^2$, where σ_i are the singular values of G.

For this, note that

$$\|NG^H(GG^H)^{-1}\|_F^2$$
$$= tr\{[NG^H(GG^H)^{-1}]^H[NG^H(GG^H)^{-1}]\}$$
$$= tr\{[(GG^H)^{-1}]^H GN^H NG^H(GG^H)^{-1}\} \quad (34)$$

Taking the expectation, we obtain

$$E\{\|NG^H(GG^H)^{-1}\|_F^2\}$$
$$= tr\{M\sigma^2[(GG^H)^{-1}]^H GG^H(GG^H)^{-1}\}$$
$$= M\sigma^2 \cdot tr\{[(GG^H)^{-1}]^H\} = M\sigma^2 \cdot tr\{(GG^H)^{-1}\} \quad (35)$$

where we have employed the earlier assumption of whiteness and uncorrelatedness of noise into different antennas so that $E\{N^H N\} = M\sigma^2 I$ (where I denotes an identity matrix). Now $tr\{(GG^H)^{-1}\}$ is equal to the sum of the eigenvalues of $(GG^H)^{-1}$, which are the inverses of the eigenvalues of GG^H or equivalently the inverses of the squared magnitudes of the singular values of G. Therefore, minimisation of the mean-square estimation error is equivalent to minimisation of the normalised error norm.