Training Sequence and Memory Length Selection for Space-Time Viterbi Equalization

Chih-Sheng Chou and David W. Lin

Abstract: **We consider signal and receiver design for space-time Viterbi equalization for wireless transmission. We propose a search method to find good training sequences, termed min-norm training sequences, for least-square channel estimation. Compared to either a maximum-length sequence or a randomly generated training sequence, the training sequence obtained can drastically reduce the channel estimation error. We also derive a simple lower bound on the achievable channel estimation error of any training sequence. Knowledge of this lower bound helps the search for min-norm training sequences in that it facilitates a measure of the goodness of the best sequence examined so far. For operation under the situation with unknown channel response lengths, we propose a simple method to select the memory length (tap number) in the Viterbi equalizer based on the SNR of the received signal. The resulting equalization performance is found to be comparable with the case where a preset, fixed memory length is used. However, the proposed method often results in use of a smaller tap number, which translates into a reduction in the computational complexity. Simulation results show that at symbol error rate below 10^{-2} (SNR > 5 dB) the amount of complexity reduction is of the order of 5% to 25% on the average, for typical wireless channels.**

Index Terms: **Wireless communication, space-time signal processing, Viterbi equalization, decision-feedback sequence estimation, channel estimation, training sequence design, channel length selection.**

I. INTRODUCTION

Wireless communication systems are evolving towards higher data rates, which experience more adverse channel effects than at the lower rates of earlier systems. On the other hand, space-time signal processing is known to be able to improve the wireless transmission performance and has attracted much recent attention [1], [2]. In this paper, we consider signal and receiver design for space-time Viterbi equalization in high-speed mobile communication. For simplicity, we consider nonspread-spectrum transmission, although the fundamental architecture could be augmented to address spread-spectrum transmission.

Fig. 1 illustrates the structure of the transmission system, where the space-time Viterbi equalizer is composed of a vector channel estimator and a vector-channel Viterbi sequence estimator. The system employs a training sequence for the equalizer's use, but the receiver does not have to know the received signal's DOA (direction of arrival) or the array manifold vector [1], [3] (also known as the array propagation vector [4]). The adaptability of the antenna array consists in the Viterbi equalizer's ability to adapt its channel model through the channel estimation unit.

First, we consider the design of the training sequence for channel estimation. A search method which combines exhaustive search and a Newton-like algorithm for finding the training sequence is proposed. By this search method, we can get a reasonably good training sequence. A simple lower bound on the achievable channel estimation error of any training sequence is also derived. Secondly, we consider using dynamic tap number selection in the Viterbi equalizer according to the channel SNR condition. Simulation results show that the approach can offer significant reduction in the receiver complexity while not sacrificing transmission performance.

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

$$\begin{aligned}\underline{x}(t) &= \sum_{i=-\infty}^{\infty} a_i[p(t-iT) * \underline{b}(t) * u(t)] + \underline{n}(t)\\ &\triangleq \sum_{i=-\infty}^{\infty} a_i \underline{r}(t-iT) + \underline{n}(t),\end{aligned} \tag{1}$$

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. Upon sampling of $\underline{x}(t)$, we obtain

$$\underline{x}(k) = R\underline{a}(k) + \underline{n}(k), \tag{2}$$

where

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

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

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

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

Consider tentatively the transmission of the training sequence only. Let the training sequence have length L. Let q be the length of the sampled vector impulse response $\underline{r}(kT)$, that is, q is the number of columns in R. Then the center section of the received data, where there is full convolution between the

Manuscript received July 28, 2000.

The authors are with the Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan, e-mail: cschou@mail2000.com.tw.

This work was supported by the National Science Council of R.O.C. under grant NSC 89-2219-E-009-013.

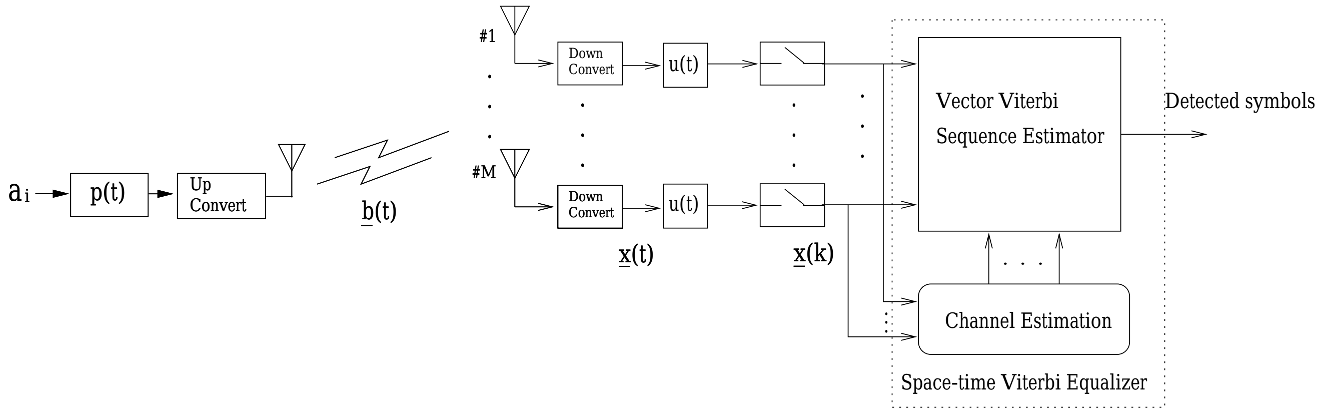

Fig. 1. A wireless transmission system with space-time Viterbi equalization.

overall channel impulse response and the training sequence, has length $L-q+1$. And we may express it in a matrix form as [1]

$$X = RG + N = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,L-q+1} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,L-q+1} \\ \vdots & \vdots & \cdots & \vdots \\ x_{M,1} & x_{M,2} & \cdots & x_{M,L-q+1} \end{bmatrix}, \quad (5)$$

where G is the training data matrix given by

$$G = [\underline{g}_1', \underline{g}_2', \cdots, \underline{g}_q']' \quad (6)$$

with

$$\underline{g}_i = [a_{q-i+1}, a_{q-i+2}, \cdots, a_{L-i+1}], \quad (7)$$

and N is a matrix of noise samples. We assume the noise into the different antennas to be zero-mean white Gaussian, uncorrelated, and with an identical variance σ^2.

In what follows, Section II characterizes the training sequence for least-square channel estimation. It also derives a simple lower bound on the estimation error for any training sequence. It then proposes a search method to find training sequences yielding the minimum estimation error (called the min-norm training sequences). Section III presents an algorithm to dynamically select the memory length (tap number) in the Viterbi equalizer according to the SNR condition. Section IV presents some simulation results and Section V gives the conclusion.

II. MIN-NORM TRAINING SEQUENCES

Let $\hat{R}$ denote the channel response estimate. Given the received data X and the training data matrix G, consider least-square estimate of the channel response as

$$\min_{\hat{R}} \|X - \hat{R}G\|_F^2, \quad (8)$$

where $\|\cdot\|_F$ denotes the Frobenius norm [10]. Simple calculus yields the optimal $\hat{R}$ as

$$\hat{R} = XG^H(GG^H)^{-1}, \quad (9)$$

where the superscript H denotes Hermitian transpose. Then the sum-square error in channel estimation is

$$\begin{aligned} \|\hat{R} - R\|_F^2 &= \|(RG+N)G^H(GG^H)^{-1} - R\|_F^2 \\ &= \|NG^H(GG^H)^{-1}\|_F^2. \end{aligned} \quad (10)$$

We now show that minimization of the mean-square estimation error is equivalent to minimization of the following quantity:

$$\sum_{i=1}^{q} \frac{1}{|\sigma_i|^2}, \quad (11)$$

where σ_i are the singular values of G.

For this, note that the least-square estimation error is given by

$$\begin{aligned} \|\hat{R} - R\|_F^2 &= \|NG^H(GG^H)^{-1}\|_F^2 \\ &= tr\{[NG^H(GG^H)^{-1}]^H[NG^H(GG^H)^{-1}]\} \\ &= tr\{[(GG^H)^{-1}]^H GN^H NG^H(GG^H)^{-1}\}. \end{aligned} \quad (12)$$

Taking the expectation, we get

$$\begin{aligned} E\{\|\hat{R} - R\|_F^2\} &= tr\{M\sigma^2[(GG^H)^{-1}]^H GG^H(GG^H)^{-1}\} \\ &= M\sigma^2 \cdot tr\{[(GG^H)^{-1}]^H\} \\ &= M\sigma^2 \cdot tr\{(GG^H)^{-1}\}, \end{aligned} \quad (13)$$

since $E\{N^H N\} = M\sigma^2 I$ (where I denotes an identity matrix) by the earlier uncorrelated AWGN assumption. Therefore, the mean-square estimation error is proportional to the sum of the eigenvalues of $(GG^H)^{-1}$. But since the eigenvalues of $(GG^H)^{-1}$ are equal to the inverses of the eigenvalues of GG^H, or equivalently the inverses of the squared magnitudes of the singular values of G, minimization of the mean-square estimation error requires minimization of the sum-square inverse singular value (11) as claimed above.

For convenience, term the sum (11) as the *normalized error norm* and term the training sequences yielding the minimum normalized error norm as *min-norm training sequences*. We have the following result.

Theorem 1: Let each symbol in the training sequence have unit magnitude. For generality, let the training data matrix G

have n rows and ℓ columns (where in the earlier discussion we had $n = q$ and $\ell = L - q + 1$). Then the normalized error norm (11) is lower bounded by n/ℓ.

Proof: Since each symbol in the training sequence has unit magnitude and the dimension of G is $n \times \ell$, the diagonal elements of GG^H are all equal to ℓ and $tr\{GG^H\} = n\ell$. Now since the trace of a matrix is equal to the sum of its eigenvalues, we have

$$tr(GG^H) = \sum_{i=1}^{n} |\sigma_i|^2 = n\ell. \quad (14)$$

If $|\sigma_i|$ were independent variables, then we could formulate the following problem:

$$\min_{\{|\sigma_i|\}} \sum_{i=1}^{n} \frac{1}{|\sigma_i|^2} \quad (15)$$

$$\text{s.t.} \sum_{i=1}^{n} |\sigma_i|^2 = n\ell. \quad (16)$$

To solve this problem, we employ the Lagrange-multiplier method to convert it into an unconstrained minimization problem as

$$y \triangleq \sum_{i=1}^{n} \frac{1}{|\sigma_i|^2} + \lambda \cdot \sum_{i=1}^{n} |\sigma_i|^2 \quad (17)$$

where λ is the Lagrange multiplier. Taking the partial derivatives of y with respect to $|\sigma_i|$, $i = 1, \cdots, n$, and setting the results equal to 0, we get

$$\frac{\partial y}{\partial |\sigma_i|} = \frac{-2}{|\sigma_i|^3} + 2\lambda |\sigma_i| = 0 \ \ \forall i. \quad (18)$$

Hence

$$|\sigma_1|^4 = |\sigma_2|^4 = \cdots = |\sigma_n|^4 = \frac{1}{\lambda}. \quad (19)$$

From (16) and (19), therefore,

$$|\sigma_1|^2 = |\sigma_2|^2 = \cdots = |\sigma_n|^2 = \ell. \quad (20)$$

And thus

$$\sum_{i=1}^{n} \frac{1}{|\sigma_i|^2} = \frac{n}{\ell}. \quad (21)$$

If G is square, that is, if $n = \ell$, then the rhs is equal to unity. □

Since G is Toeplitz in actual systems, $|\sigma_i|$ are not independent variables as assumed in the foregoing proof. For common modulation schemes (BPSK, QPSK, etc.), furthermore, they are not even continuous variables. Therefore, the lower bound may not be reachable in actual systems. The reachable minimum value under these conditions is yet unknown and subject to further study. It is also interesting to note that the foregoing results on optimal training sequences in the uncorrelated AWGN environment do not depend on the number of antennas. Therefore, an optimal training sequence is optimal for an array-based receiver regardless of the number of antenna branches in the array, as long as the length of channel response stays the same.

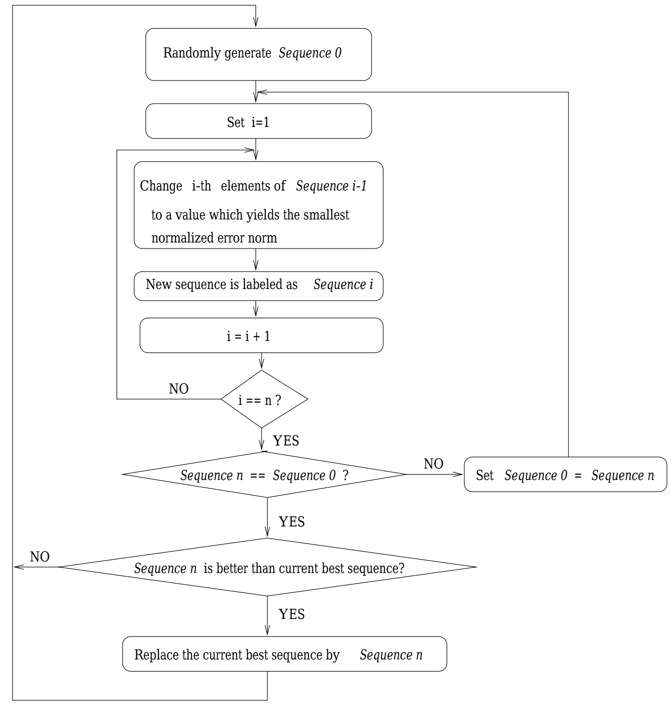

Fig. 2. Flowchart explaining the search procedure for min-norm sequence.

A. Search for Min-Norm Training Sequences

Since no simple constructive methods exist for obtaining the min-norm training sequences, we resort to a search approach to find sequences with small normalized error norms. However, exhaustive search is often impossible. In the simulation study reported later, for example, we make use of a training sequence of 29 QPSK symbols. In this case, there are 4^{28} possible sequences! We therefore employ Newton's method with multiple initial points. The search method generates an initial data sequence at random and modifies the sequence in a way to move the point toward a direction of descent till it reaches the bottom of a valley. Since the result may be a local minimum, the procedure is repeated a number of times with a number of random initial points, in hope to find the global minimum or a better local minimum. Because the normalized error norm has a lower bound (equal to 1 when the training data matrix is square), a possible stopping criterion for the search is when we have obtained a sequence whose associated normalized error norm is within a certain tolerance of the lower bound. A disadvantage of this method is that we cannot predict the required search time. But this is not a serious drawback, since the search is conducted in the stage of system design and not in the stage of its actual operation.

Since the training sequence is composed of discrete numbers, the Newton steps are somewhat different from that in the case of continuous numbers. The search procedure is illustrated in Fig. 2. We summarize the search procedure as follows:

1. Randomly generate a sequence and call it Sequence 0.
2. For $i = 1$ to n where n is the length of the training sequence, do the following: Change the i-th element of Sequence $i - 1$ to a value which yields the smallest normal-

ized error norm, i.e., $tr\{(GG^H)^{-1}\}$. Call the resulting sequence Sequence i.

3. Check if Sequence n is the same as Sequence 0. If so, then compare the associated normalized error norm with that of the current best sequence; replace the current best sequence by Sequence n if the latter has a lower error norm; go to Step 1. Otherwise, rename Sequence n as Sequence 0 and go to Step 2.

We conducted a search over QPSK sequences of length 29 and found that the following sequence yielded the lowest normalized error norm when the sequence is formed into a 15×15 square training data matrix:

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

The value of the normalized error norm is 1.058, which is very close to the lower bound of 1. In fact, it can be shown that, since the training data matrix is of an odd dimension, it is impossible to have a sequence with unity normalized error norm under either BPSK or QPSK. We state the previous result in the form of a theorem below. As a result, the above sequence may be one of the best sequences of length 29 under QPSK modulation.

Theorem 2: When the training data matrix is square and of an odd dimension, it is impossible to have a sequence with unity normalized error norm under either BPSK or QPSK.

Proof: Consider eigenvalue decomposition of the matrix GG^H as EVE^H where V is the diagonal matrix of the eigenvalues of GG^H and E is the corresponding matrix of eigenvectors. The proof of Theorem 1 shows that, if G achieves the lower bound of normalized error norm, then all the eigenvalues of GG^H are equal (and they are equal to the column dimension of G). In the case of a square G of dimension n, therefore, we would have $GG^H = EVE^H = nEE^H = nI$, which implies that the rows of G are orthogonal. But such orthogonality is impossible between any BPSK or QPSK sequences of odd lengths, because an odd number of ± 1 (in the case of BPSK) or an odd number of ± 1 and $\pm j$ (in the case of QPSK) cannot sum to zero. Thus the corresponding sequence cannot have unity normalized error norm. □

An interesting question is whether some of the well-known kinds of sequences in communications, such as the maximum-length (ML) sequences [6], would be optimal or nearly optimal. While a theoretical derivation is not conducted in this work, the simulation results presented later show that ML sequences can yield significantly higher channel estimation error than a sequence which minimizes the normalized 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. The simulation results presented later also show that a random training sequence can be much worse, as expected.

Incidentally, an independent study on training sequence design was reported by Crozier *et al.* [12]. Among other differences, they focused on single-antenna reception under (primarily) BPSK modulation. In addition, our derivation of the min-norm condition is more direct and succinct.

III. DYNAMIC SELECTION OF VITERBI EQUALIZER'S MEMORY LENGTH

Intuitively, the memory length (number of taps in the channel response model) of the Viterbi equalizer should vary with the length of the channel response. However, previous study shows that when the receiver input SNR is low, use of fewer taps in the Viterbi equalizer may yield a better performance than using more. This is because the truncation of some small-valued taps can reduce the channel estimation error when the SNR is low [11]. Moreover, use of fewer taps in Viterbi equalizer can also reduce the computational complexity. We now propose a simple method to determine the memory length of the Viterbi equalizer, under the condition that the input SNR is known. By selecting memory length this way, we can achieve the same performance as using a larger number of taps.

Consider again the received training signal (5) and the least-square estimate of the channel response (9). From (9) we have

$$\hat{R} = X \cdot G^H (GG^H)^{-1} = R + N \cdot G^H (GG^H)^{-1}. \quad (22)$$

Considering the use of QPSK modulation, then the input SNR can be defined as

$$SNR = \frac{\|R\|_F^2}{E\{\|N\|_F^2\}}. \quad (23)$$

Let the SNR in $\hat{R}$, denoted $S\hat{N}R$, be defined as

$$S\hat{N}R = \frac{\|R\|_F^2}{E\{\|N\|_F^2\}\|Z\|_F^2} = \frac{SNR}{\|Z\|_F^2}, \quad (24)$$

where $Z = G^H (GG^H)^{-1}$.

Assume the original channel contains q taps, but only the k-th to the $(k+L_{win}-1)$-th taps are selected in the Viterbi equalizer; that is, only L_{win} taps are used in the Viterbi equalizer. From (22) we can find that the noise variances in all the estimated taps are the same; this is due to that each column of Z has the same sum-squares value for its elements. Therefore, we estimate the noise energy in each estimated tap, denoted E_n, as

$$E_n = \frac{\hat{P}}{(S\hat{N}R + 1) \cdot q}, \quad (25)$$

where $\hat{P} \triangleq \|\hat{R}\|_F^2$ is the energy of the estimated channel $\hat{R}$. The total estimation error over the length of channel modeled in the Viterbi equalizer can be estimated as

$$E_s = L_{win} \cdot E_n. \quad (26)$$

And the channel truncation error can be estimated as

$$E_t = (\hat{P} - \hat{P}_{win}) - (q - L_{win}) \cdot E_n, \quad (27)$$

where $\hat{P}_{win}$ is the energy in the selected taps of the estimated channel $\hat{R}$.

The total channel estimation error can thus be estimated as

$$E_s + E_t = 2 \cdot L_{win} \cdot E_n + (\hat{P} - \hat{P}_{win}) - q \cdot E_n. \quad (28)$$

In the rhs, since $\hat{P}$ and $q \cdot E_n$ do not vary with L_{win}, we may choose L_{win} by performing the following minimization:

$$\min_{L_{win}} \left\{ 2 \cdot L_{win} \cdot \frac{\hat{P}}{(S\hat{N}R+1) \cdot q} - \hat{P}_{win} \right\}. \qquad (29)$$

When SNR is small, the cost to be minimized in (29) is dominated by the first term; that is, fewer taps will result a smaller error. On the other hand, when SNR is large, the cost is dominated by $\hat{P}_{win}$. In this condition, using a larger number of taps is worthwhile. As to which taps are selected for a given L_{win}, we pick the L_{win} consecutive taps that result in the maximum sum-square value in the estimated channel $\hat{R}$. From (24), the estimation of SNR also requires estimation of $E\{\|N\|_F^2\}$ and $\|R\|_F^2$. In a practical receiver, the former can be achieved by noise power estimation in the absence of signal, while the latter can be achieved by observing that, from (22), $\|R\|_F^2 = \|\hat{R}\|_F^2 - \sigma^2 \cdot tr\{(GG^H)^{-1}\}$. In the simulation reported below, we assume for simplicity that $\|R\|_F^2$ and the noise power are known exactly.

IV. SIMULATION RESULTS

In the simulated space-time Viterbi equalizer, we employ the delayed-decision sequence estimation architecture [8] wherein the channel response estimate is provided by a channel response estimator. To evaluate the performance of various techniques, 100 channel responses were generated according to the model of [7]. Then the average performance of these techniques over this set of channels was obtained. The model of [7] covers different propagation environments, including urban, suburban, and hilly. We generated channels with rms delay spread between about 1 and 10 μs. The QPSK signals ($a_i \in \{\pm 1, \pm j\}$) were transmitted at 1 Mbaud with raised-cosine pulse shaping with $\alpha = 0.75$. The receiver employs three antennas. We assumed perfect carrier and timing recovery. To facilitate computation in our simulation, the channel impulse response matrix R is truncated to contain 99.9% of the power in the overall channel response $\underline{r}(kT)$ (see Section I).

Firstly, consider the performance of the min-norm training sequence. We compare it with that of a ML sequence and a randomly generated training sequence, where the random training sequence is considered for reference purpose. The number of taps covered in the maximum-likelihood sequence estimator (MLSE) part is 4 and that in the DFE (decision-feedback equalizer) part is 1. Fig. 3 shows the symbol error rates from using different kinds of training sequences. The performance of the min-norm training sequence is about 1.5 dB better than that of the ML sequence. Both are much better than the performance of the randomly generated training sequence.

Next, consider the performance of dynamical selection of the memory length for the Viterbi equalizer. In Fig. 4, the symbol error rates from using dynamic tap number selection and a fixed tap number are shown. In the latter case, 4 taps are used in the MLSE part and 1 tap for the DFE part, while in the former, the possible tap numbers are 3 and 4 for the MLSE part and 0 and 1 for the DFE part. We see that the performance of these two schemes is near the same.

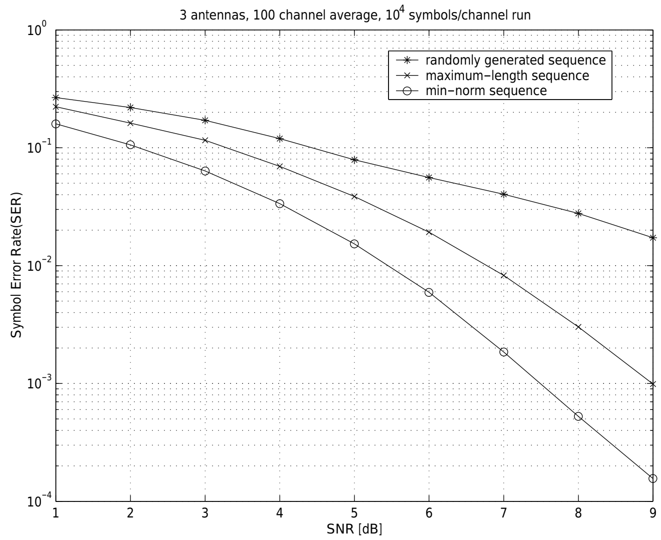

Fig. 3. Symbol error rates from using different training sequences. The tap numbers in the MLSE and the DFE parts are 4 and 1, respectively.

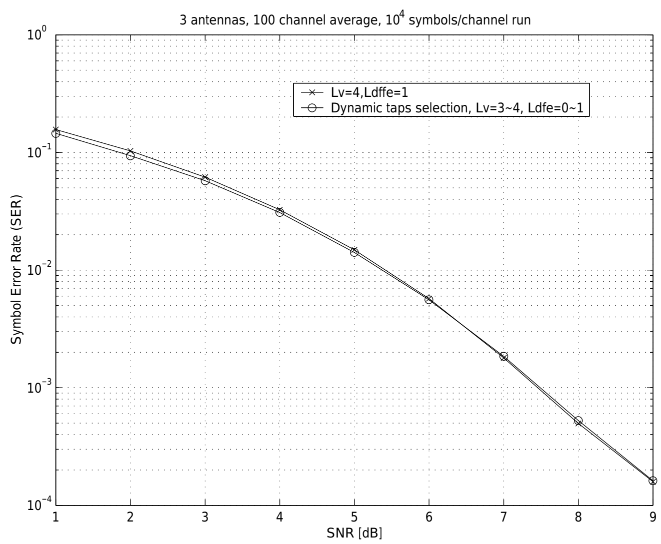

Fig. 4. Symbol error rates from using dynamic tap number selection and a fixed tap number.

Finally, consider the comparative computational complexity between using a fixed tap number and using dynamic tap number selection. We use L_v to denote the tap number in the MLSE part of the equalizer, and L_{dfe} denotes the tap number in the DFE part of the equalizer. And L_{win} is equal to the sum of L_v and L_{dfe}. We use 4^{L_v-1} to give a rough estimate of the complexity for each channel, since this value gives the number of metrics computed in MLSE. In the case with a fixed tap number, $L_v = 4$ because the number of taps in the MLSE part is 4. In the case with dynamic tap number selection, L_v is either 3 or 4 depending on the channel condition. The ratio in computational complexity of these two schemes, averaged over the 100 test channels for each simulated SNR value, is shown in Fig. 5. We see that at symbol error rates below 10^{-2} (SNR > 5 dB), dynamic tap number selection can reduce the compu-

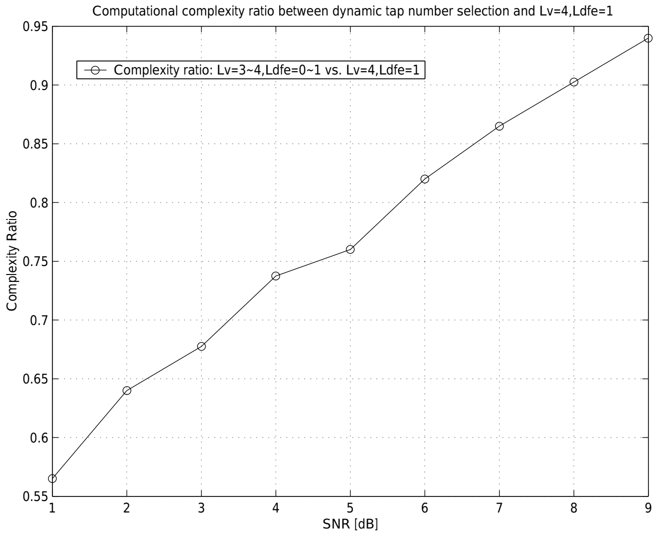

Fig. 5. Ratio in computational complexity between using dynamic tap number selection and a fixed tap number.

tational complexity by about 5% to 25% compared to using a fixed tap number, while achieving the same performance.

Note that, with dynamic memory-length Viterbi equalization, the hardware complexity is still determined by the largest allowed memory length. But the reduced computational complexity will result in reduced power consumption in system operation.

V. CONCLUSION

We considered signal and receiver design for space-time Viterbi equalization in non-spread-spectrum mobile communication at high data rates, where the Viterbi equalizer is of the delayed-decision sequence estimation variant. The examined system structure employed a training sequence for channel estimation and used the estimate in space-time Viterbi equalization.

We proposed a search method to find the best training sequences (called the min-norm training sequences) for least-square channel estimation. Simulation results show that such training sequences can achieve a much better performance in channel estimation and signal transmission than using a maximum-length sequence or a randomly generated sequence in training. We also derived a simple lower bound for the achievable channel estimation error of any training sequence. One interesting aspect of the theoretical results is that the optimality of a training sequence does not depend on the number of antenna branches used in the receiver. Interpreted another way, the receiver may freely choose the number of antennas to use without concern for the optimality of the training sequence. This is certainly beneficial to system design.

For the situation with unknown channel response lengths, we proposed a simple method to select the memory length of the Viterbi equalizer based on the receiver input SNR. Simulation results show that the performance of the Viterbi equalizer under this method is comparable to that using a fixed memory length, even though the memory length (tap number) in the former case is often shorter than that in the latter. On the average, at symbol error rates below 10^{-2} (SNR > 5 dB), the reduced tap number translates into a reduction of about 5% to 25% in computational complexity of the Viterbi equalizer.

Least-square channel estimation method requires matrix inversions, in principle. When the dimension of the matrices or the number of such inversions is high, the computational complexity can be a concern in practical implementation. However, since the matrices that need be inverted are solely a function of the training sequence and hence are known in advance, they can be pre-computed and stored. Thus matrix inversions do not constitute a part of the per-sample computational complexity, but only some matrix multiplications.

REFERENCES

[1] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.
[2] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, New York: Wiley, 1980.
[3] T.-S. Lee, "Efficient wideband source localization using beamforming invariance techniques," *IEEE Trans. Signal Processing*, vol. 42, pp. 1376–1387, June 1994.
[4] J. Litva and T. K.-Y. Lo, *Digital Beamforming in Wireless Communications*, Boston: Artech House, 1996.
[5] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.
[6] D. V. Sarwate and M. B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," *Proc. IEEE*, vol. 68, no. 5, pp. 593–619, May 1980.
[7] H. Hashemi, "Simulation of the urban radio propagation channel," *IEEE Trans. Veh. Technol.*, vol. 28, no. 3, pp. 213–225, Aug. 1979.
[8] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.
[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1980.
[10] P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Application*, 2nd Ed., Academic Press, 1985.
[11] C.-S. Chou and D. W. Lin, "Training techniques of space-time Viterbi equalizers for high speed mobile communication," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2000, pp. II-89–II-92.
[12] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proc. F,*, vol. 138, no. 4, pp. 371–378, Aug. 1991.

Chih-Sheng Chou was born in Pingtung, Taiwan, in 1970. He received the B.S. and M.S. degree in electrical engineering from Tatung University, Taipei, Taiwan, in 1992 and 1994, respectively. He is currently working towards the Ph.D. degree at National Chiao-Tung University, Hsinchu, Taiwan. His research interests include signal processing and wireless communications.

David W. Lin received the B.S. degree from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1975, and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1979 and 1981, respectively, all in electrical engineering.

He was with Bell Laboratories during 1981–1983, and with Bellcore during 1984–1990 and again during 1993–1994. Since 1990, he has been a Professor in the Department of Electronics Engineering and the Center for Telecommunications Research, National Chiao Tung University. He has conducted research in digital adaptive filtering and telephone echo cancellation, digital subscriber line and coaxial network transmission, speech and video coding, and wireless communication. His research interests include various topics in communication engineering and signal processing.