Symbol-Timing Estimation in Space–Time Coding Systems Based on Orthogonal Training Sequences

Yik-Chung Wu, S. C. Chan, *Member, IEEE*, and Erchin Serpedin, *Senior Member, IEEE*

***Abstract*—Space–time coding has received considerable interest recently as a simple transmit diversity technique for improving the capacity and data rate of a channel without bandwidth expansion. Most research in space–time coding, however, assumes that the symbol timing at the receiver is perfectly known. In practice, this has to be estimated with high accuracy. In this paper, a new symbol-timing estimator for space–time coding systems is proposed. It improves the conventional algorithm of Naguib *et al.* such that accurate timing estimates can be obtained even if the oversampling ratio is small. Analytical mean-square error (MSE) expressions are derived for the proposed estimator. Simulation and analytical results show that for a modest oversampling ratio (such as Q equal to four), the MSE of the proposed estimator is significantly smaller than that of the conventional algorithm. The effects of the number of transmit and receive antennas, the oversampling ratio, and the length of training sequence on the MSE are also examined.**

***Index Terms*—Approximated log-likelihood function, space–time coding, symbol-timing recovery, training sequences.**

I. Introduction

SPACE–TIME (ST) processing using ST coding has received considerable interest recently as an efficient means for high-rate data transmission [1]–[10]. Symbol-timing synchronization is an important issue in ST coding systems because perfect symbol-timing information at the receiver is usually assumed. This problem was first studied in [4], where orthogonal training sequences are transmitted at different transmit antennas to simplify the maximization of the oversampled approximated log-likelihood function. The sample having the largest magnitude, the so-called the "optimal sample," is assumed to be closest to the optimum sampling instants (it will be referred to as the optimum sample selection algorithm in the sequel for convenience). However, it is shown in this paper that the mean square error (MSE) of this algorithm is lower bounded by $1/(12Q^2)$, where Q is the oversampling ratio. As a result, the performance of this timing synchronization method depends highly on the oversampling ratio. In fact, relatively high oversampling ratios might be required for accurate symbol-timing estimation.

In this paper, a new symbol-timing estimator for ST coding systems is proposed. It improves the optimum sample selection algorithm in [4] so that accurate timing estimates can be obtained even if the oversampling ratio is small. The increase in implementation complexity with respect to that of optimum sample selection algorithm is very small. The requirements and the design procedures for the training sequences are discussed. Analytical expressions for MSE of the proposed estimator are derived. Both analytical and simulation results show that, for a modest oversampling ratio (such as $Q = 4$), the MSE of the proposed estimator is significantly smaller than that of the optimum sample selection algorithm. Furthermore, the effects of the number of transmit and receive antennas, the oversampling ratio, and the length of training sequence on the MSE are also examined.

The paper is organized as follows. The system model of the ST coding system is first described in Section II. A brief overview of the optimum sample selection algorithm for symbol-timing synchronization in an ST coding system is given in Section III. Requirements and design of training sequences are discussed in Section IV. The proposed symbol-timing estimator is then presented in Section V. Analytical MSE expressions are derived in Section VI. Simulation results and discussions are then presented in Section VII, and finally conclusions are drawn in Section VIII.

II. Signal Model

Both ST block coding and ST trellis coding systems can be described by the same basic communication model [4]. The simplified baseband equivalent model, with N transmit and M receive antennas, is shown in Fig. 1. The information is encoded by an ST trellis or block encoder to give the encoded symbols $d_1(l), d_2(l), \ldots, d_N(l)$. Each encoded data symbol $d_i(l)$ is pulse shaped and then transmitted simultaneously via different antennas. A superposition of independently faded signals from all the transmit antennas plus noise is received at each receive antenna. The received signal at each receive antenna is first filtered by a matched filter. It is then passed through the symbol-timing recovery and the channel estimation units and, finally, the ST decoder.

For a flat fading channel, the received signal at the jth receive antenna can be written as

$$r_j(t) = \sqrt{\frac{E_s}{N}} \sum_{i=1}^{N} h_{ij} \sum_{n} d_i(n) g(t - nT - \epsilon T) + n_j(t), \quad j = 1, 2, \ldots, M \quad (1)$$

Manuscript received March 31, 2003; revised August 14, 2003; accepted January 20, 2004. The editor coordinating the review of this paper and approving it for publication is C. Xiao. This work was supported by the National Science Foundation under Award CCR-0092901.

Y.-C. Wu was with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China. He is currently with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843-3128 USA (e-mail: ycwu@ee.tamu.edu).

S. C. Chan is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China.

E. Serpedin is with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843-3128 USA (e-mail: serpedin@ee.tamu.edu).

Digital Object Identifier 10.1109/TWC.2004.843063

1536-1276/$20.00 © 2005 IEEE

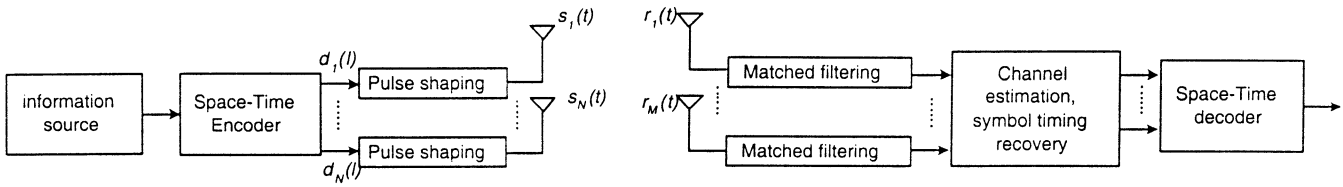

Fig. 1. Simplified baseband equivalent model for ST coding system.

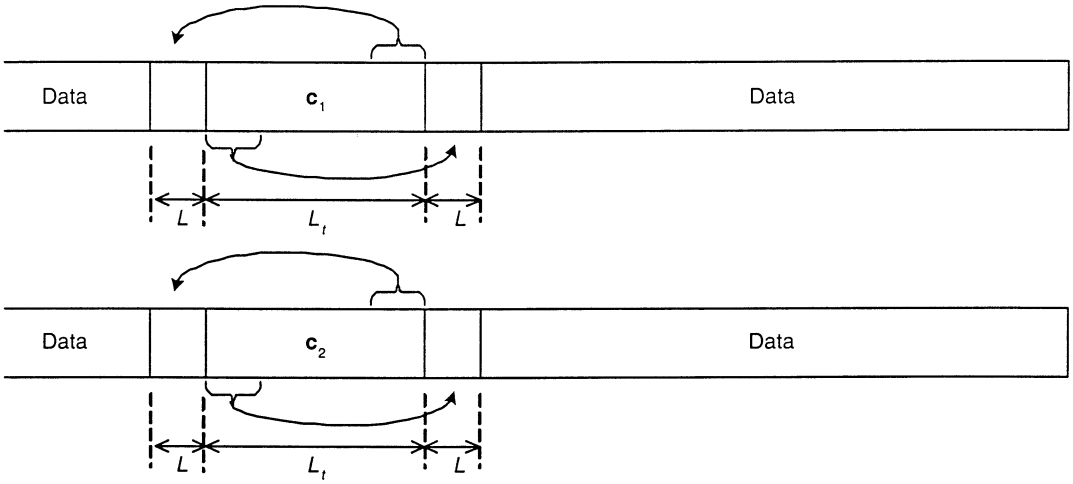

Fig. 2. Structure of the training sequence for symbol-timing synchronization in a two-transmit antenna system.

where E_s/N is the symbol energy and h_{ij}s are the complex channel coefficients between the ith transmit antenna and the jth receive antenna and are assumed to be statistically independent for different transmit/receive antennas (this condition is satisfied if the transmit antennas are well separated, e.g., more than $\lambda/2$, where λ is the wavelength of the RF carrier). $d_i(n)$ is the information symbol transmitted from the ith transmit antenna; $g(t)$ is the transmit filter, which is assumed to be a root raised cosine pulse; T is the symbol duration; $\epsilon \in [-0.5, 0.5]$ is the unknown timing offset; and $n_j(t)$ is the complex-valued circularly distributed Gaussian white noise at the jth receive antenna, with power density N_o. Throughout this paper, it is assumed that the channel is frequency flat and quasi-static.

Let the received signal be sampled at a rate Q times faster than the symbol rate $1/T$. The sampled and matched filtered signal at the jth receive antenna is given by

$$r_j(m) = \sqrt{\frac{E_s}{N}} \sum_{i=1}^{N} h_{ij} \sum_{n} d_i(n) p(mT/Q - nT - \epsilon T) + \eta_j(m) \quad (2)$$

where[1] $r_j(m) \triangleq r_j(mT/Q)$, $p(t) \triangleq g(t) \otimes g_r(t)$, $\eta_j(m) \triangleq n_j(t) \otimes g_r(t)|_{t=mT/Q}$, and $g_r(t)$ denotes the matched filter.

III. Timing Synchronization by Optimum Samples Selection

As proposed in [4], orthogonal training sequences can be periodically transmitted in between data symbols (as shown in Fig. 2) to assist the timing synchronization. The idea is that at the receiver, if the position of the orthogonal training sequences can be correctly located, the signal from any one of the transmit antennas can be extracted (and signals from other antennas are removed) by multiplying the received signal with the orthogonal sequence transmitted from that antenna. Note that the structure of training sequences in this paper is different from that presented in [4]. In this paper, a cyclic prefix and cyclic suffix, each of length L, are included in order to remove the intersymbol interference (ISI) from the random data transmitted before and after the orthogonal training sequences. Since L is usually kept as a small number, the increase in length of training is very small, especially when the length of the orthogonal training sequences is large.

Let $\mathbf{c}_i = [c_i(0)\ c_i(1)\ \cdots\ c_i(L_t - 1)]$ be the ith ($i = 1, \ldots, N$) orthogonal training sequence of length L_t to be transmitted from the ith transmit antenna. The sampled signal at the jth receive antenna can be obtained by replacing $d_i(n)$ in (2) with $c_i(n)$. Further, let $m = lQ + k (l = 0, 1, \ldots, L_t - 1$ and $k = k_o, k_o + 1, \ldots, k_o + Q - 1$, where $k_o = -\lfloor (1/2 - \epsilon)Q \rfloor$ and $\lfloor x \rfloor$ denotes the nearest integer less than or equal to x) so that each sample is indexed by the lth training bit and the kth phase. In order to maintain the orthogonality between the received training sequences and the local copies, the first phase is taken at $-\lfloor (1/2 - \epsilon)Q \rfloor$ such that all the Q samples for the lth training bit are taken from $-T/2 \le t - lT \le T/2$. Then, the received signal $r_j(lQ + k)$, due to the orthogonal training sequences, can be rewritten as

$$r_j(lQ + k) = \sqrt{\frac{E_s}{N}} \sum_{i=1}^{N} h_{ij} \sum_{n} c_i(n) \times p(kT/Q + (l - n)T - \epsilon' T) + \eta_j(lQ + k) \quad \text{for } l = 0, 1, \ldots, L_t - 1,$$

and

$$k = 0, 1, \ldots, Q - 1 \quad (3)$$

[1]Notation $\triangleq$ stands for "is defined as," and $\otimes$ denotes convolution.

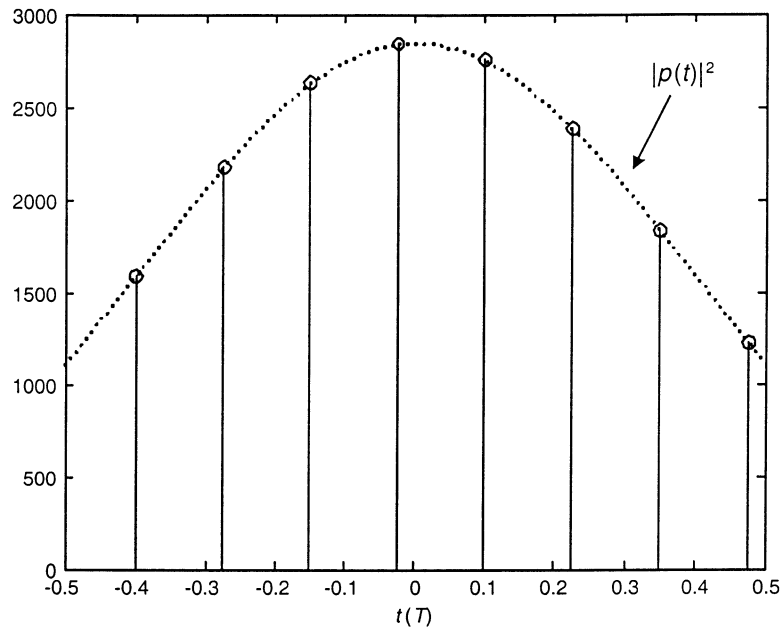

Fig. 3. $\Lambda_{ij}(k)$ with the scaled version of $|p(t)|^2$ for $-T/2 \le t \le T/2$ (dotted line).

where $\epsilon' \triangleq \epsilon + k_o/Q$. Note that k_o has been dropped from the index of $\eta_j(lQ+k)$ since a fixed time shift does not affect the noise statistics. In practice, it is sufficient to estimate ϵ' only as it represents the time difference between the first sample of the training sequence and the next nearest optimum sampling instance. Grouping the samples with the same phase, one can form the vector $\mathbf{r}_j(k)$ as in (4) and (5), shown at the bottom of the page. Please also refer to the other equation shown at the bottom of the page.[2]

Define the sequence $\Psi_{ij}(k) \triangleq \mathbf{c}_i^H \mathbf{r}_j(k)$, where $\mathbf{c}_i^H$ denotes the transpose conjugate of $\mathbf{c}_i$. Since $\mathbf{c}_i$s are orthogonal to each other when the relative delay is zero, it follows that

$$\Psi_{ij}(k) = \sqrt{\frac{E_s}{N}} h_{ij} p\left(\frac{kT}{Q} - \epsilon' T\right) \|\mathbf{c}_i\|^2 + \sqrt{\frac{E_s}{N}} \sum_{i'=1}^{N} h_{i'j} \mathbf{c}_i^H \tilde{\mathbf{C}}_{i'} \tilde{\mathbf{p}}(k) + \mathbf{c}_i^H \boldsymbol{\eta}_j(k) \quad (6)$$

for $k = 0, \ldots, Q-1$, $\|\mathbf{c}_i\| \triangleq \sqrt{\mathbf{c}_i^H \mathbf{c}_i}$ is the norm of $\mathbf{c}_i$, which is a constant. $\tilde{\mathbf{C}}_i$ is the same as $\mathbf{C}_i$ but with the $(L+1)$th column removed, and $\tilde{\mathbf{p}}(k)$ is the same as $\mathbf{p}(k)$ but with the $(L+1)$th entry removed. The second term in (6) represents the ISI if the training sequences are not orthogonal when the relative delay is not zero. The last term in (6) is the noise term.

From (6), it can be observed that if the second and third terms are very small [a training sequence design procedure that makes the second term zero is discussed in the next section; the third term is small at high signal-to-noise ratios (SNRs)], $\Psi_{ij}(k)$ has the same shape as $p(t)$ for $-T/2 \le t \le T/2$, except that it is scaled by a complex channel gain and is corrupted by additive noise. In order to remove the effect of the channel, consider the sequence $\Lambda_{ij}(k) \triangleq |\Psi_{ij}(k)|^2$. Now, the sequence $\Lambda_{ij}(k)$ should have a similar shape to the function $|p(t)|^2$ for $-T/2 \le t \le T/2$. This is illustrated in Fig. 3, where an example sequence of $\Lambda_{ij}(k)$ is shown ($Q = 8, L_t = 32, L = 3$ and in the absence of noise). Note that a scaled version of $|p(t)|^2$ for $-T/2 \le t \le T/2$ is also shown (in dotted line) for comparison. It can be seen that the optimum sampling time is at $t = 0$ and the sample with maximum amplitude is the one closer to the optimum sampling instant than the remaining samples.

A simple symbol-timing synchronization algorithm is to choose a value of k closest to the optimum sampling instants. That is, the optimum sampling phase $k = \hat{k}$ is selected such that

[2]Notation $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$.

$$\mathbf{r}_j(k) \triangleq [r_j(k) \quad r_j(Q+k) \quad r_j(2Q+k) \quad \cdots \quad r_j((L_t-1)Q+k)]^T \quad (4)$$

$$= \sqrt{\frac{E_s}{N}} \sum_{i=1}^{N} h_{ij} \mathbf{C}_i \mathbf{p}(k) + \boldsymbol{\eta}_j(k) \quad (5)$$

$$\mathbf{C}_i \triangleq \begin{bmatrix} c_i(\mathrm{mod}(-L, L_t)) & c_i(\mathrm{mod}(-L+1, L_t)) & \cdots & c_i(\mathrm{mod}(L, L_t)) \\ c_i(\mathrm{mod}(-L+1, L_t)) & c_i(\mathrm{mod}(-L+2, L_t)) & \cdots & c_i(\mathrm{mod}(L+1, L_t)) \\ \vdots & \vdots & & \vdots \\ c_i(\mathrm{mod}(-L+L_t-1, L_t)) & c_i(\mathrm{mod}(-L+L_t, L_t)) & \cdots & c_i(\mathrm{mod}(L+L_t-1, L_t)) \end{bmatrix}$$

$$\mathbf{p}(k) \triangleq [p(kT/Q - LT - \epsilon' T) \quad p(kT/Q - (L-1)T - \epsilon' T) \quad \cdots \quad p(kT/Q + LT - \epsilon' T)]^T$$

$$\boldsymbol{\eta}_j(k) \triangleq [\eta_j(k) \eta_j(Q+k) \quad \cdots \quad \eta_j((L_t-1)Q+k)]^T$$

it maximizes $\Lambda_{ij}(k)$. Since the channels for different antennas are independent, the average of $\Lambda_{ij}(k)$ over all i and j is maximized (see (8), where the scaling factor $1/MN$ is not included in order to preserve a simplified notation). As mentioned in [4], this represents the samples of an approximated log-likelihood function for symbol-timing synchronization, when the ISI plus noise term in (6) is assumed to be Gaussian. Therefore, the optimum sampling phase is selected as [4]

$$\hat{k} = \max_{k=0,1,\ldots,Q-1} \Lambda_{ML}(k) \tag{7}$$

with

$$\Lambda_{ML}(k) = \sum_{j=1}^{M}\sum_{i=1}^{N} \Lambda_{ij}(k). \tag{8}$$

Under the optimistic assumption that the samples closest to the optimum sampling positions are correctly estimated (at high SNRs), the estimation error, normalized with respect to the symbol duration, is a uniformly distributed random variable in the range $[-1/2Q, 1/2Q]$. Therefore, the MSE is $1/(12Q^2)$. Thus, a relatively high oversampling ratio might be required in order to obtain a small MSE.

IV. Design of Training Sequences

The performance of the proposed timing estimator is directly influenced by the presence of ISI and noise terms in (6). In order to minimize the contribution of the ISI term in (6), the training sequences need to be designed such that

$$\mathbf{c}_i^H \tilde{\mathbf{C}}_{i'} = 0 \tag{9}$$

for all combinations of i and i'. Combining with the fact that sequences from different antennas have to be orthogonal when the relative delay is zero, the problem of training sequences design resumes to finding N sequences such that

$$\mathbf{C}_i^H \mathbf{C}_{i'} = \begin{cases} \|\mathbf{c}_i\|^2 \mathbf{I}, & \text{if } i = i' \\ 0, & \text{if } i \neq i' \end{cases}$$

where $\mathbf{I}$ denotes the identity matrix. This is exactly the problem of designing multiple $(2L+1)$-perfect sequences [11]–[13], with each of length L_t. Here, we just mention the procedures for designing the training sequences, interested readers can refer to the original papers [11]–[13] for details.

1) Construct a sequence $\mathbf{s} = [s(0)\ s(1)] \cdots\ s(L_t - 1)]$ with length L_t such that all of its out-of-phase periodic auto-correction terms are equal to zero. One example of this kind of sequence is a Chu sequence [14].
2) Construct another sequence $\mathbf{s}' = [s'(0)\ s'(1)\ \cdots\ s'(L_t + 2NL - 1)]$ of length $L_t + 2NL$ as in (10), shown at the bottom of the page. Note that $L_t \geq 2NL$ must be satisfied. That is, if the number of transmit antennas N is large, we cannot use training sequences with short length.
3) The orthogonal training sequences are given by

$$\mathbf{c}_i = [s'((2i-1)L) \quad \cdots \quad s'((2i-1)L + L_t - 1)]. \tag{11}$$

For example, let us consider $L_t = 32, L = 3, N = 2$. First, we construct a Chu sequence of length 32. Then, we cyclically extend the Chu sequence by copying the first $2 \times 2 \times 3 = 12$ bits and putting them at the back. Then, $\mathbf{c}_1 = [s'(3)\ s'(4)\ \cdots\ s'(34)]$ and $\mathbf{c}_2 = [s'(9)\ s'(10)\ \cdots\ s'(40)]$.

V. Timing Synchronization by Estimation

In an optimum samples selection algorithm, symbol timing is estimated by maximization of the oversampled approximated log-likelihood function. As the number of samples becomes very large (which requires a large oversampling ratio), the estimate could become accurate. However, noting that the approximated log likelihood function is "smooth" (see Fig. 3), we expect that the maximization of the log-likelihood function can be done by interpolation based on a few samples, thus keeping the oversampling ratio at a small number.

More precisely, let us construct a periodic sequence $\tilde{\Lambda}_{ML}(m)$ by periodically extending the approximated log-likelihood sequence $\Lambda_{ML}(k)$ in (8). Further, denote $\tilde{\Lambda}_{ML}(\hat{\epsilon}')$ as the continuous and periodic approximated log-likelihood function with its samples given by $\tilde{\Lambda}_{ML}(m)$. According to the sampling theorem, as long as the sampling frequency Q/T is higher than twice the highest frequency of $\tilde{\Lambda}_{ML}(\hat{\epsilon}')$, then $\tilde{\Lambda}_{ML}(\hat{\epsilon}')$ can be represented by its samples $\tilde{\Lambda}_{ML}(m)$ without loss of information. The relationship between $\tilde{\Lambda}_{ML}(\hat{\epsilon}')$ and $\tilde{\Lambda}_{ML}(m)$ is then given by

$$\tilde{\Lambda}_{ML}(\hat{\epsilon}') = \sum_{m=-\infty}^{\infty} \tilde{\Lambda}_{ML}(m) \mathrm{sinc}\left(\pi \frac{\hat{\epsilon}'T - mT/Q}{T/Q}\right). \tag{12}$$

Now, expand $\tilde{\Lambda}_{ML}(\hat{\epsilon}')$ into a Fourier series

$$\tilde{\Lambda}_{ML}(\hat{\epsilon}') = \sum_{\ell=-\infty}^{\infty} A_\ell e^{j2\pi\ell\hat{\epsilon}'} \tag{13}$$

where

$$A_\ell = \int_0^1 \tilde{\Lambda}_{ML}(\hat{\epsilon}') e^{-j2\pi\ell\hat{\epsilon}'} d\hat{\epsilon}'. \tag{14}$$

Substituting (12) into (14) yields

$$\begin{aligned} A_\ell &= \sum_{m=-\infty}^{\infty} \tilde{\Lambda}_{ML}(m) \\ &\quad \times \int_0^1 \mathrm{sinc}\left(\pi \frac{\hat{\epsilon}'T - mT/Q}{T/Q}\right) e^{-j2\pi\ell\hat{\epsilon}'} d\hat{\epsilon}' \\ &= \sum_{k=0}^{Q-1} \Lambda_{ML}(k) \\ &\quad \times \sum_{l=-\infty}^{\infty} \int_0^1 \mathrm{sinc}\left(\pi \frac{\hat{\epsilon}'T - lT - kT/Q}{T/Q}\right) e^{-j2\pi\ell\hat{\epsilon}'} d\hat{\epsilon}' \\ &= \sum_{k=0}^{Q-1} \Lambda_{ML}(k) e^{-j2\pi\ell k/Q} \frac{1}{Q} \mathfrak{F}\{\mathrm{sinc}(\pi\hat{\epsilon}')\}_{f=\ell/Q} \end{aligned} \tag{15}$$

$$\mathbf{s}' = [\underbrace{s(0) \quad s(1) \quad \cdots \quad s(L_t - 1)}_{\mathbf{s}} s(0) \quad s(1) \quad \cdots \quad s(2NL - 1)] \tag{10}$$

where $\mathfrak{F}\{\}$ denotes the Fourier transform. It is clear that if Q is even, we have the equation shown at the bottom of the page, and if Q is odd, we have the other equation shown at the bottom of the page.

From (13), it can be seen that once the coefficients A_ℓ are determined, the timing delay ϵ' can be estimated by maximizing $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ for $0 \leq \hat{\epsilon'} \leq 1$. Note that $\Lambda_{ML}(k)$ only contains samples of the approximated log-likelihood function at certain delays, while $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ is a continuous function of $\hat{\epsilon'}$. Therefore, maximizing $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ provides a more accurate estimate of the timing delay than maximizing $\Lambda_{ML}(k)$. For efficient implementation, the maximization can be performed by discrete Fourier transform (DFT)-based interpolation. More specifically, $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ for $0 \leq \hat{\epsilon'} \leq 1$ can be approximated by a K-point sequence, denoted as $\Lambda_{ML}(k')$ for $0 \leq k' \leq K-1$, by zero padding the high frequency coefficients of A_ℓ and performing a K-point inverse discrete Fourier transform (IDFT). For a sufficiently large value of $K, \Lambda_{ML}(k')$ becomes very close to $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ for $0 \leq \hat{\epsilon'} \leq 1$, and the index with the maximum amplitude can be viewed as an improved estimate of the timing parameter ϵ'.

To avoid the complexity in performing the K-point IDFT, an approximation is applied to (13). More precisely, extensive simulations show that $A_{\pm 1}$ are much greater than A_ℓ for $|\ell| > 1$; therefore,

$$\tilde{\Lambda}_{ML}(\hat{\epsilon'}) \approx A_0 + 2\mathfrak{Re}\{A_1 e^{j2\pi\hat{\epsilon'}}\} \quad \text{for } 0 \leq \hat{\epsilon'} \leq 1 \tag{16}$$

where $\mathfrak{Re}\{x\}$ stands for the real part of x. In order to maximize the approximated log-likelihood function $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$, we have

$$\arg(A_1) = -2\pi\hat{\epsilon'} \tag{17}$$

where $\arg(x)$ denotes the phase of x. Equivalently

$$\hat{\epsilon'} = -\frac{1}{2\pi}\arg\left\{\sum_{k=0}^{Q-1}\Lambda_{ML}(k)e^{-j2\pi k/Q}\right\}. \tag{18}$$

The estimated delay $\hat{\epsilon'}$ is the time between the first sampling phase and the nearest optimum sampling instant. The calculation within the arg-operation is actually the second output of a Q-point DFT of the sequence (or the Fourier coefficient at symbol rate $f = 1/T$). Note that the increase in complexity of the proposed algorithm in (18) with respect to that of optimum samples selection algorithm is only a Q-point DFT (which can be efficiently implemented using Goertzel's algorithm) and an arg-operation. From the simulation results to be presented at Section VII, it is found that an oversampling factor Q of four is sufficient to yield good estimates in practical applications. Therefore, the four-point DFT in (18) can be computed easily without any multiplications since $\exp(-j2\pi k/4) \in \{\pm 1, \pm j\}$. This greatly reduces the arithmetic complexity of implementation.

VI. Performance Analysis

We derive the MSE expressions of the proposed estimator as a function of E_s/N_o in this section. First, express the true delay as

$$\epsilon' = -\frac{1}{2\pi}\arg(e^{-j2\pi\epsilon'}). \tag{19}$$

Taking the difference between (18) and (19), the MSE is given by

$$\mathbf{E}[(\hat{\epsilon'} - \epsilon')^2] = \left(\frac{1}{2\pi}\right)^2 \times \mathbf{E}\left[\left(\arctan\left\{\frac{\mathfrak{Im}\phi}{\mathfrak{Re}\phi}\right\}\right)^2\right] \tag{20}$$

where

$$\phi \triangleq e^{j2\pi\epsilon'}\sum_{k=0}^{Q-1}\Lambda_{ML}(k)e^{-j2\pi k/Q}. \tag{21}$$

Applying the approximation $\arctan(x) \approx x$ for small x, we have

$$\begin{aligned}
&\mathbf{E}[(\hat{\epsilon'} - \epsilon')^2] \\
&\approx \left(\frac{1}{2\pi}\right)^2 \mathbf{E}\left[\left(\frac{\phi - \phi^*}{j(\phi + \phi^*)}\right)^2\right] && (22)\\
&\approx -\left(\frac{1}{2\pi}\right)^2 \frac{\mathbf{E}[\phi^2] - 2\mathbf{E}[\phi\phi^*] + \mathbf{E}[(\phi^*)^2]}{\mathbf{E}[\phi^2] + 2\mathbf{E}[\phi\phi^*] + \mathbf{E}[(\phi^*)^2]} && (23)\\
&= -\left(\frac{1}{2\pi}\right)^2 \frac{\mathfrak{Re}\{\mathbf{E}[\phi^2]\} - \mathbf{E}[\phi\phi^*]}{\mathfrak{Re}\{\mathbf{E}[\phi^2]\} + \mathbf{E}[\phi\phi^*]}. && (24)
\end{aligned}$$

The second approximation is justified by the fact that the mean of the denominator $\mathbf{E}[(\phi + \phi^*)^2]$ is much larger than the mean of the numerator $\mathbf{E}[(\phi - \phi^*)^2]$ (which is illustrated in Fig. 4 for $\epsilon = -0.5, -0.25, 0, 0.25$, and 0.5 with $N = 2, M = 4$, $\alpha = 0.3, Q = 4, L_t = 32$, and $L = 4$), and the variance of the numerator and denominator are much smaller than the mean of the denominator (which is true for medium to high SNRs). Some additional explanations regarding this approximation can be found in [18]. From (23) and (24), we have the fact $\mathbf{E}[(\phi^*)^2] = (\mathbf{E}[\phi^2])^*$ was used.

It is proved in Appendix I that

$$\begin{aligned}
\mathbf{E}[\phi^2] = E_s^2 L_t^2 MN e^{j4\pi\epsilon'}\Bigg[&L_t^2\frac{(1+MN)}{N^2}\Xi_{SS}^2 \\
&+ \left(\frac{E_s}{N_o}\right)^{-1}\frac{2L_t}{N}\Xi_{SN} + \left(\frac{E_s}{N_o}\right)^{-2}\Xi_{NN}\Bigg]
\end{aligned} \tag{25}$$

$$A_\ell = \begin{cases} \frac{1}{Q}\sum_{k=0}^{Q-1}\Lambda_{ML}(k)e^{-j2\pi\ell k/Q}, & \text{for } \ell = -Q/2+1, \ldots, Q/2-1 \\ \frac{1}{2Q}\sum_{k=0}^{Q-1}\Lambda_{ML}(k)e^{-j2\pi\ell k/Q}, & \text{for } \ell = -Q/2, Q/2 \\ 0, & \text{otherwise} \end{cases}$$

$$A_\ell = \begin{cases} \frac{1}{Q}\sum_{k=0}^{Q-1}\Lambda_{ML}(k)e^{-j2\pi\ell k/Q}, & \text{for } \ell = -\lfloor Q/2 \rfloor, \ldots, \lfloor Q/2 \rfloor \\ 0, & \text{otherwise} \end{cases}$$

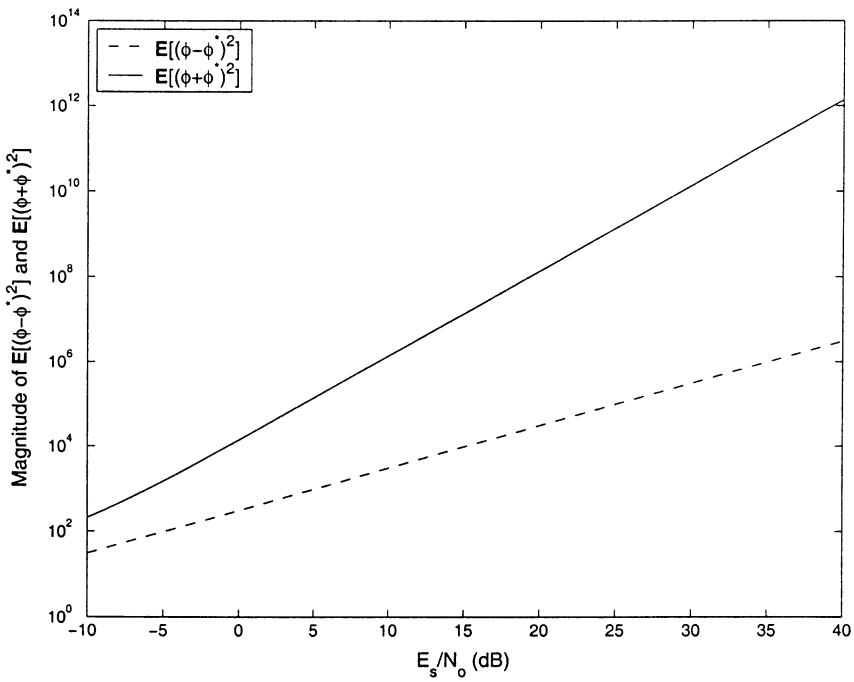

Fig. 4. Magnitude of $\mathbf{E}[(\phi-\phi^*)^2]$ and $\mathbf{E}[(\phi+\phi^*)^2]$ as a function of E_s/N_o for $\epsilon = -0.5, -0.25, 0, 0.25$, and $0.5 (N = 2, M = 4, \alpha = 0.3, Q = 4, L_t = 32, L = 4)$. Note that all curves for different values of ϵ overlap.

$$\mathbf{E}[\phi\phi^*] = E_s^2 L_t^2 MN \left[L_t^2 \frac{(1+MN)}{N^2} |\Xi_{SS}|^2 + \left(\frac{E_s}{N_o}\right)^{-1} \frac{2L_t}{N} \tilde{\Xi}_{SN} + \left(\frac{E_s}{N_o}\right)^{-2} \tilde{\tilde{\Xi}}_{NN} \right] \quad (26)$$

where

$$\Xi_{SS} \triangleq \sum_{k=0}^{Q-1} p^2(kT/Q - \epsilon' T) e^{-j2\pi k/Q} \quad (27)$$

$$\Xi_{SN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} p(k'T/Q - \epsilon' T) p(k''T/Q - \epsilon' T) \times \varphi((k'-k'')T/Q) e^{-j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (28)$$

$$\Xi_{NN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} \varphi^2((k'-k'')T/Q) e^{-j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (29)$$

$$\tilde{\Xi}_{SN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} p(k'T/Q - \epsilon' T) p(k''T/Q - \epsilon' T) \times \varphi((k'-k'')T/Q) e^{j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (30)$$

$$\tilde{\tilde{\Xi}}_{NN} \triangleq \sum_{k'=0}^{Q-1} \sum_{k''=0}^{Q-1} \varphi^2((k'-k'')T/Q) e^{j2\pi k''/Q} e^{-j2\pi k'/Q} \quad (31)$$

and

$$\varphi(\tau) \triangleq \int_{-\infty}^{\infty} g_r(t) g_r^*(t+\tau)\, dt \quad (32)$$

is the correlation between noise samples introduced by the matched filter.

Since the timing delay is assumed to be uniformly distributed, the average MSE can be calculated by numerical integration of (24).

VII. Simulation Results and Discussions

The performances of the synchronizers based on the optimum sample selection (7) and the proposed algorithm (18) are evaluated in this section. The MSE of the estimates are calculated using both the analytic expressions derived in the last section and Monte Carlo simulations, where each MSE value is obtained by averaging over 10^5 estimates. The timing offset ϵ is generated to be uniformly distributed in the interval $[-0.5, 0.5]$. The channel coefficients h_{ij} are generated as complex Gaussian random variables with zero mean and a variance of 0.5 per dimension. The raised cosine pulse with excess bandwidth $\alpha = 0.3$ is considered. The training sequences are generated following the procedures in Section IV with $L = 4$. In all the figures, the MSE of both the proposed algorithm and the optimum sample selection algorithm are plotted against E_s/N_o, with the markers showing the simulation results while the solid lines represent the theoretical MSE derived in the last section.

A. Effect of Oversampling Ratio

In (12), it is assumed that the sampling frequency is at least twice the highest frequency of $\tilde{\Lambda}_{ML}(\hat{\epsilon}')$. Since $\tilde{\Lambda}_{ML}(\hat{\epsilon}')$ has the same shape as $|p(t)|^2$ for $-T/2 \le t \le T/2$, where $p(t)$ is a raised cosine pulse, it is natural to predict that the sampling frequency Q/T has to be greater than $2\times 2/T$ (i.e., $Q \ge 4$). This prediction is corroborated by Fig. 5, where the MSE are shown for $Q = 2, 4, 8$, and 16 in a two-transmit four-receive antenna system with $L_t = 32$. Several conclusions can be drawn from the figure.

1) Performances of the optimum sample selection algorithm are lower bounded by $1/(12Q^2)$ and are poorer than that of the proposed algorithm, for all values of Q

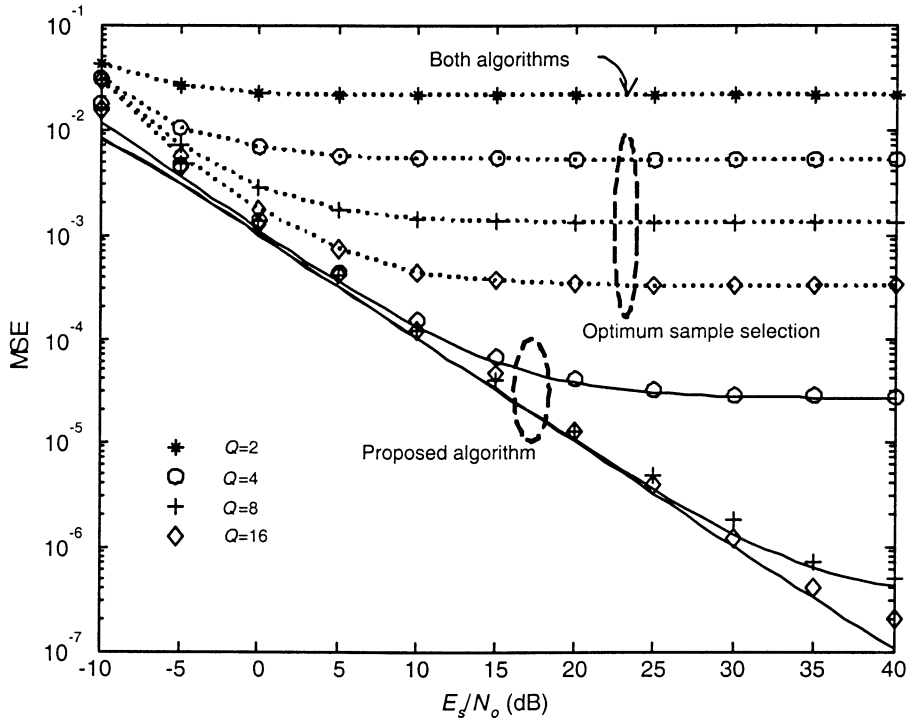

Fig. 5. MSE performance for different oversampling ratios $Q(N = 2, M = 4, L_t = 32, \alpha = 0.3)$.

(except $Q = 2$, in which case the performances of both algorithms are the same).

2) It can be seen that for $Q = 2$, the MSE of the proposed algorithm is much higher than that corresponding to other oversampling ratios. This confirms the above argument that Q has to be greater than four in order to represent $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ without much loss of information. For $Q = 8$ and 16 the performance improves at high E_s/N_o. This can be explained by the fact that $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ is a truncated version of $|p(t)|^2$, so $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ is no longer bandlimited. Therefore, $\tilde{\Lambda}_{ML}(m)$ would, in general, suffer from aliasing from the neighboring spectra. Increasing Q thus reduces the aliasing and improves the performance.

3) The analytical MSEs (solid lines in the figure) match very well with the simulation results for $Q = 4, 8$, and 16. Note that for $Q = 2$, the analytic MSE expression does not hold and only the simulation results have been plotted in Fig. 5.

4) Strictly speaking, Q should be at least equal to 16 in order to represent $\tilde{\Lambda}_{ML}(\hat{\epsilon'})$ using its samples $\tilde{\Lambda}_{ML}(m)$ without loss of information. However, for $Q = 4$, the MSE of the proposed algorithm reaches the order of 10^{-5} at medium and high E_s/N_o, which is a reasonably good performance in practical applications. Because of this reason, $Q = 4$ is used to generate the simulation results for the rest of this paper.

B. Effect of Length of Training Sequences

Fig. 6 shows the MSE of a two-transmit four-receive antenna system with different lengths (L_t) of the training sequences. In this figure, it can be seen that increasing the length of training sequences improves the performance at low E_s/N_o. But at high E_s/N_o, the MSEs are the same for all L_t. Again, the performance of the proposed algorithm is much better than that of optimum samples selection algorithm. It is also notable that the analytic MSE expressions match the simulation results very well.

C. Effect of Number of Receive Antennas

Fig. 7 compares the MSE for different numbers of receive antennas when two-transmit antennas and $L_t = 32$ are used. We can see that increasing the number of receive antennas reduces the MSE at low E_s/N_o, but it does not help at high E_s/N_o. The proposed algorithm exhibits much smaller MSE than the optimum sample selection algorithm. When comparing the theoretical and simulation results of the proposed algorithm, it can be seen that they match pretty well except for the $M = 1$ case. This is due to the fact that the $\arctan$ approximation in (22), in general, holds only for AWGN channels[3] but not for fading channels. In the presence of fading, the channel output may assume a large range of values and the approximation does not hold anymore. Of course, a better approximation, such as $\arctan(x) \approx x - x^3/3 + x^5/5$ may be used, but the analysis would become extremely complicated as higher order moments are involved. Fortunately, as the number of transmit or receive antennas increases, the equivalent averaged channel across all transmit/receive antennas tends to behave like an AWGN channel and the approximation becomes valid again. This can be seen from the cases $M = 2$ and $M = 4$, the theoretical and the simulation results are closer when compared with the $M = 1$ case. For $M = 8$ and $M = 16$, the theoretical and the simulation results match exactly.

[3]Note that this approximation has been applied in similar applications [16], [17] in AWGN channels only.

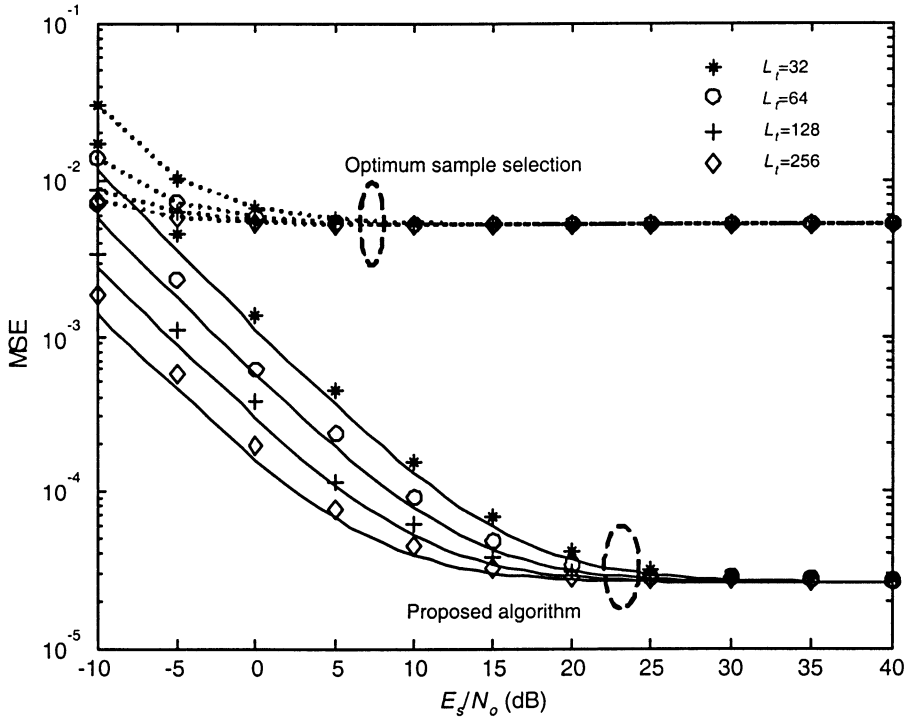

Fig. 6. MSE performance for different lengths of the training sequence ($N = 2, M = 4, Q = 4, \alpha = 0.3$).

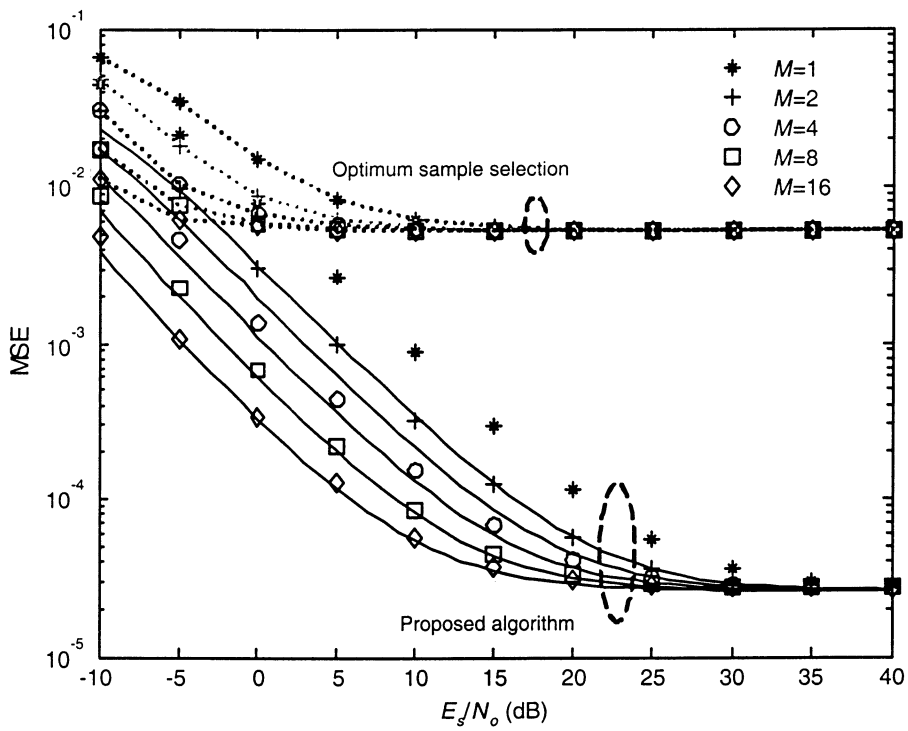

Fig. 7. MSE performance for different numbers of receive antennas $M(N = 2, L_t = 32, Q = 4, \alpha >= 0.3)$.

D. Effect of Number of Transmit Antennas

Finally, we assess the MSE when different numbers of transmit antennas are used with $L_t = 64$. The results shown in Fig. 8 illustrate that increasing the number of transmit antennas does not change the MSE performance. The theoretical and the simulation results for the proposed algorithm match very well. Once again, the proposed algorithm performs much better.

VIII. Conclusion

A new symbol-timing delay estimator for ST coding systems has been proposed. It improves the optimum sample selection algorithm of Naguib *et al.* [4] such that accurate timing estimates are obtained even if the oversampling ratio is small. The increase in implementation complexity with respect to the optimum sample selection algorithm is very small. The require-

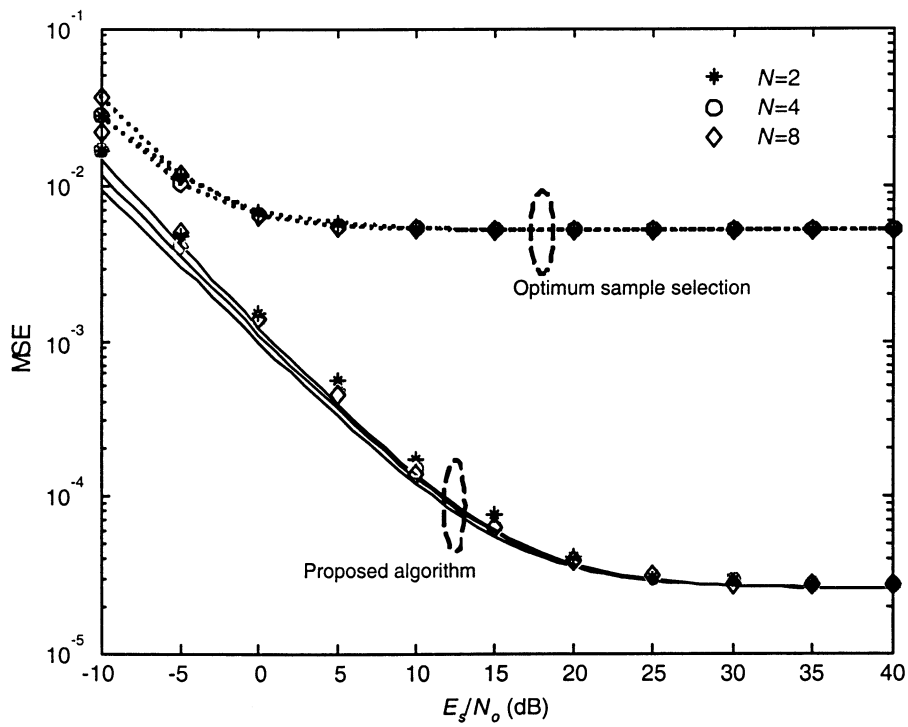

Fig. 8. MSE performance for different numbers of transmit antennas $N(L_t = 64, Q = 4, \alpha = 0.3, M = 2)$.

ments and the design procedure for the training sequences are discussed. Analytical expressions for MSE of the proposed estimator are derived. It is shown that the MSE analytical expressions match very well with the simulation results in most of the cases. Simulation results also show that for modest oversampling ratios (such as $Q = 4$), the MSE of the proposed estimator is significantly smaller than that of the optimum sample selection algorithm. Furthermore, the performance of the proposed algorithm improves with the number of receive antennas being employed or the length of training sequences.

Appendix I
Calculation of $\mathbf{E}[\phi^2]$ and $\mathbf{E}[\phi\phi^*]$

Since we can construct orthogonal sequences such that (9) is satisfied, the ISI term in (6) vanishes. Further, with the fact that $\|\mathbf{c}_i\|^2 = L_t$, we have

$$\begin{aligned}\Lambda_{ML}(k) &= \sum_{i=1}^{N}\sum_{j=1}^{M}|\Psi_{ij}(k)|^2 \\ &= W^2 h p^2(kT/Q - \epsilon' T) + v(k)\end{aligned} \tag{33}$$

where

$$W \triangleq \sqrt{E_s/N}L_t \tag{34}$$

$$h \triangleq \sum_{i=1}^{N}\sum_{j=1}^{M}|h_{ij}|^2 \tag{35}$$

$$\begin{aligned}v(k) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{M}\Big\{ &\left|\mathbf{c}_i^H \boldsymbol{\eta}_j(k)\right|^2 \\ &+ 2Wp(kT/Q - \epsilon' T)\mathfrak{Re}\left[h_{ij}\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k)\right)^*\right]\Big\}.\end{aligned} \tag{36}$$

Then, (21) can be rewritten as

$$\begin{aligned}\phi = W^2 h e^{j2\pi\epsilon'}\sum_{k=0}^{Q-1}p^2(kT/Q - \epsilon' T)e^{-j2\pi k/Q} \\ + e^{j2\pi\epsilon'}\sum_{k=0}^{Q-1}v(k)e^{-j2\pi k/Q}.\end{aligned} \tag{37}$$

Before we proceed to the calculation of $\mathbf{E}[\phi^2]$ and $\mathbf{E}[\phi\phi^*]$, we first calculate the mean and the second moment of $v(k)$. Note the following facts:

$$\mathbf{E}[\eta_j(l_1Q + k')] = 0 \quad \forall j, l_1, k' \tag{38}$$

$$\mathbf{E}[h_{ij}\eta_j(l_1Q + k')] = 0 \quad \forall i, j, l_1, k' \tag{39}$$

$$\mathbf{E}[\eta_j(l_1Q + k')\eta_{j'}(l_2Q + k'')] = 0 \quad \forall j, j', l_1, l_2, k', k'' \tag{40}$$

$$\begin{aligned}&\mathbf{E}[\eta_j(l_1Q + k')\eta_{j'}^*(l_2Q + k'')] \\ &\quad = N_o\varphi(((l_1 - l_2)Q + k' - k')T/Q)\delta_{jj'}\end{aligned} \tag{41}$$

$$\mathbf{E}[h_{ij}h_{i'j'}] = 0 \quad \forall i, i', j, j' \tag{42}$$

$$\mathbf{E}[h_{ij}h_{i'j'}^*] = \delta_{ii'}\delta_{jj'} \tag{43}$$

where $\delta_{ii'} = 1$ if $i = i'$ and zero otherwise. Since the matched filter is a root raised cosine filter, we also have

$$\varphi(0) = 1 \tag{44}$$

$$\varphi(\tau) = \varphi(-\tau) \tag{45}$$

$$\varphi(lT) = 0 \quad \text{for } l \neq 0. \tag{46}$$

Let

$$\begin{aligned}v_{ij}(k) = \left|\mathbf{c}_i^H\boldsymbol{\eta}_j(k)\right|^2 + 2Wp(kT/Q - \epsilon' T) \\ \times\mathfrak{Re}\left[h_{ij}\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k)\right)^*\right]\end{aligned} \tag{47}$$

such that $v(k) = \sum_{i=1}^{N}\sum_{j=1}^{M} v_{ij}(k)$. The mean of $v_{ij}(k)$ is

$$\begin{aligned}\mathbf{E}[v_{ij}(k)] &= \sum_{l=0}^{L_t-1}\sum_{l'=0}^{L_t-1} c_i(l')c_i^*(l)\mathbf{E}[\eta_j(lQ+k)\eta_j^*(l'Q+k)] \\ &= \sum_{l=0}^{L_t-1} |c_i(l)|^2\mathbf{E}[|\eta_j(lQ+k)|^2] \\ &= N_oL_t \end{aligned} \tag{48}$$

where in the first equality, we applied (39) and in the second equality, we applied (46). Therefore, $\mathbf{E}[v(k)] = MNN_oL_t$ is a constant and independent of k.

The second moment of $v_{ij}(k)$ is given by

$$\begin{aligned}&\mathbf{E}[v_{ij}(k')v_{i'j'}(k'')] \\ &\quad= \mathbf{E}\left[\left|\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right|^2\left|\mathbf{c}_{i'}^H\boldsymbol{\eta}_{j'}(k'')\right|^2\right] \\ &\qquad+ 4W^2p(k'T/Q-\epsilon'T)p(k''T/Q-\epsilon'T) \\ &\qquad\cdot\mathbf{E}\left\{\mathfrak{Re}\left[h_{ij}\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right)^*\right]\right. \\ &\qquad\left.\times\,\mathfrak{Re}\left[h_{i'j'}\left(\mathbf{c}_{i'}^H\boldsymbol{\eta}_{j'}(k'')\right)^*\right]\right\}. \end{aligned} \tag{49}$$

Note that (39) makes the cross terms vanish. Considering the first term in (49)

$$\begin{aligned}&\mathbf{E}\left[\left|\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right|^2\left|\mathbf{c}_{i'}^H\boldsymbol{\eta}_{j'}(k'')\right|^2\right] \\ &\quad= \sum_{l_1=0}^{L_t-1}\sum_{l_2=0}^{L_t-1}\sum_{l_3=0}^{L_t-1}\sum_{l_4=0}^{L_t-1} c_i^*(l_1)c_i(l_2)c_{i'}^*(l_3)c_{i'}(l_4) \\ &\qquad\cdot\mathbf{E}[\eta_j(l_1Q+k')\eta_j^*(l_2Q+k') \\ &\qquad\times\eta_{j'}(l_3Q+k'')\eta_j^*(l_4Q+k'')]. \end{aligned} \tag{50}$$

Using the fact that if a, b, c, d are jointly Gaussian, then

$$\begin{aligned}\mathbf{E}[abcd] &= \mathbf{E}[ab]\mathbf{E}[cd] + \mathbf{E}[ac]\mathbf{E}[bd] \\ &\quad+ \mathbf{E}[ad]\mathbf{E}[bc] + \mathbf{E}[a]\mathbf{E}[b]\mathbf{E}[c]\mathbf{E}[d] \end{aligned} \tag{51}$$

and applying (38), (40), and (41), we have

$$\begin{aligned}&\mathbf{E}[\eta_j(l_1Q+k')\eta_j^*(l_2Q+k')\eta_{j'}(l_3Q+k'')\eta_j^*(l_4Q+k'')] \\ &\quad= N_o^2[\varphi((l_1-l_2)T)\varphi((l_3-l_4)T)] \\ &\qquad+ N_o^2[\varphi((l_1-l_4)T+(k'-k'')T/Q) \\ &\qquad\times\varphi((l_2-l_3)T+(k'-k'')T/Q)]\delta_{jj'}. \end{aligned} \tag{52}$$

Plugging this result back into (50), we obtain

$$\begin{aligned}&\mathbf{E}\left[\left|\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right|^2\left|\mathbf{c}_{i'}^H\boldsymbol{\eta}_{j'}(k'')\right|^2\right] \\ &\quad= N_o^2L_t^2 + N_o^2\left\{\sum_{l_1=0}^{L_t-1}\sum_{l_4=0}^{L_t-1} c_i^*(l_1)c_{i'}(l_4)\right. \\ &\qquad\times\varphi((l_1-l_4)T+(k'-k'')T/Q) \\ &\qquad\cdot\sum_{l_2=0}^{L_t-1}\sum_{l_3=0}^{L_t-1} c_i^*(l_2)c_{i'}(l_3) \\ &\qquad\left.\times\,\varphi((l_2-l_3)T+(k'-k'')T/Q)\right\}\delta_{jj'}. \end{aligned} \tag{53}$$

Consider first $i \neq i'$. We note that the second term in (53) is approximately zero since $\varphi(\tau)$ is a decaying function of τ. When $l_1 = l_4$ or $|l_1 - l_4|$ is small, $\varphi((l_1 - l_4)T + (k' - k'')T/Q)$ has significant values. But in these cases, $\sum_{l_1=0}^{L_t-1}\sum_{l_4=0}^{L_t-1} c_i^*(l_1)c_{i'}(l_4) = 0$ since the training sequences are designed such that they are orthogonal when the relative delay is small. When $|l_1 - l_4|$ is large, $\varphi((l_1 - l_4)T + (k' - k'')T/Q) \approx 0$. The same argument applies to $\varphi((l_2 - l_3)T + (k' - k'')T/Q)$. For $i = i'$, the only case that the second term in (53) is nonzero is when $l_1 = l_4$ and $l_2 = l_3$. Therefore, we have

$$\begin{aligned}&\mathbf{E}\left[\left|\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right|^2\left|\mathbf{c}_{i'}^H\boldsymbol{\eta}_{j'}(k'')\right|^2\right] \\ &\qquad= N_o^2L_t^2\left(1+\varphi^2((k'-k'')T/Q)\delta_{ii'}\delta_{jj'}\right). \end{aligned} \tag{54}$$

Now consider the second term of (49) (ignoring the nonrandom part at this moment). Expanding it out and applying (42) and (43), we note that it is zero except for the case $i = i'$ and $j = j'$, in which case we have

$$\begin{aligned}&4\mathbf{E}\left\{\mathfrak{Re}\left[h_{ij}\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right)^*\right]\right. \\ &\quad\left.\times\,\mathfrak{Re}\left[h_{ij}\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k'')\right)^*\right]\right\} \\ &\qquad= \mathbf{E}\left[h_{ij}\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right)^* h_{ij}^*\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k'')\right)\right] \\ &\qquad\quad+ \mathbf{E}\left[h_{ij}^*\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k')\right)h_{ij}\left(\mathbf{c}_i^H\boldsymbol{\eta}_j(k'')\right)^*\right] \\ &\qquad= N_o\mathbf{E}[|h_{ij}|^2]\sum_{l_1=0}^{L_t-1}\sum_{l_2=0}^{L_t-1} c_i(l_1)c_i^*(l_2) \\ &\qquad\quad\times\varphi((l_2-l_1)T+(k''-k')T/Q) \\ &\qquad\quad+ N_o\mathbf{E}[|h_{ij}|^2]\sum_{l_1=0}^{L_t-1}\sum_{l_2=0}^{L_t-1} c_i^*(l_1)c_i(l_2) \\ &\qquad\quad\times\varphi((l_1-l_2)T+(k'-k'')T/Q) \\ &\qquad= 2N_oL_t\varphi((k'-k'')T/Q). \end{aligned} \tag{55}$$

Plugging (54) and (55) back into (49), we obtain

$$\begin{aligned}&\mathbf{E}[v_{ij}(k')v_{i'j'}(k'')] \\ &\quad= N_o^2L_t^2 + N_o^2L_t^2\varphi^2((k'-k'')T/Q)\delta_{ii'}\delta_{jj'} \\ &\qquad+ 2N_oL_tW^2p(k'T/Q-\epsilon'T)p(k''T/Q-\epsilon'T) \\ &\qquad\times\varphi((k'-k'')T/Q)\delta_{ii'}\delta_{jj'}. \end{aligned} \tag{56}$$

Finally

$$\begin{aligned}&\mathbf{E}[v(k')v(k'')] \\ &\quad= \sum_{j=1}^{M}\sum_{i=1}^{N}\mathbf{E}[v_{ij}(k')v_{ij}(k'')] \\ &\qquad+ \sum_{j=1}^{M}\sum_{i=1}^{N}\sum_{\substack{j'=1\\ j'\neq j}}^{M}\sum_{\substack{i'=1\\ i'\neq i}}^{N}\mathbf{E}[v_{ij}(k')v_{i'j'}(k'')] \\ &\quad= MN(MN+\varphi^2((k'-k'')T/Q))N_o^2L_t^2 \\ &\qquad+ 2MNN_oL_tW^2p(k'T/Q-\epsilon'T)p(k''T/Q-\epsilon'T) \\ &\qquad\times\varphi((k'-k'')T/Q). \end{aligned} \tag{57}$$

Now, return to the calculation of $\mathbf{E}[\phi^2]$ and $\mathbf{E}[\phi\phi^*]$. From (37), we have

$$\mathbf{E}[\phi^2] = W^4\mathbf{E}[h^2]e^{j4\pi\epsilon'}\left(\sum_{k=0}^{Q-1} p^2(kT/Q-\epsilon' T)e^{-j2\pi k/Q}\right)^2 + e^{j4\pi\epsilon'}\mathbf{E}\left[\left(\sum_{k=0}^{Q-1} v(k)e^{-j2\pi k/Q}\right)^2\right]. \quad (58)$$

The cross terms vanish since h and $v(k)$ are uncorrelated and $\mathbf{E}[v(k)]$ is a constant and independent of k. Note that h is a central chi-square random variable with $2MN$ degrees of freedom and the variance in each dimension equals 0.5, so $\mathbf{E}[h^2] = MN(1+MN)$. Using (57), it can be easily shown that

$$\mathbf{E}\left[\left(\sum_{k=0}^{Q-1} v(k)e^{-j2\pi k/Q}\right)^2\right] = 2MNN_oL_tW^2\Xi_{SN} + MNN_o^2L_t^2\Xi_{NN} \quad (59)$$

where

$$\Xi_{SN} \triangleq \sum_{k'=0}^{Q-1}\sum_{k''=0}^{Q-1} p(k'T/Q-\epsilon' T)p(k''T/Q-\epsilon' T) \times \varphi((k'-k'')T/Q)e^{-j2\pi k''/Q}e^{-j2\pi k'/Q} \quad (60)$$

$$\Xi_{NN} \triangleq \sum_{k'=0}^{Q-1}\sum_{k''=0}^{Q-1} \varphi^2((k'-k'')T/Q)e^{-j2\pi k''/Q}e^{-j2\pi k'/Q}. \quad (61)$$

Plugging (59) back into (58), the expression for $\mathbf{E}[\phi^2]$ can be obtained and is given by (25). A similar procedure can be applied to obtain the expression for $\mathbf{E}[\phi\phi^*]$.

REFERENCES

[1] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Process. Mag.*, vol. 17, pp. 76–92, May 2000.

[2] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.

[3] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 10, pp. 1451–1458, Oct. 1998.

[4] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 10, pp. 1459–1478, Oct. 1998.

[5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 744–765, Mar. 1998.

[6] E. G. Larsson, P. Stoica, and J. Li, "On the maximum-likelihood detection and decoding for space-time coding system," *IEEE Trans. Signal Process.*, vol. 50, pp. 937–944, Apr. 2002.

[7] H. E. Gamal, "On the robustness of space-time coding," *IEEE Trans. Signal Process.*, vol. 50, pp. 2417–2428, Oct. 2002.

[8] Z. Liu and G. B. Giannakis, "Space-time block coded multiple access through frequency selective fading channels," *IEEE Trans. Commun.*, vol. 49, no. 6, pp. 1033–1045, Jun. 2001.

[9] A. R. Hammons Jr. and H. E. Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans. Inf. Theory*, no. 3, pp. 524–542, Mar. 2000.

[10] G. Yi and K. B. Letaief, "Performance evaluation and analysis of space-time coding in unequalized multipath fading links," *IEEE Trans. Commun.*, vol. 48, no. 11, pp. 1778–1782, Nov. 2000.

[11] C. Fragouli, N. Al-Dhahir, and W. Turin, "Finite-alphabet constant-amplitude training sequence for multiple-antenna broadband transmission," in *Proc. ICC 2002*, pp. 6–10.

[12] ——, "Reduced-complexity training schemes for multiple-antenna broadband transmissions," in *Proc. WCNC 2002*, pp. 78–83.

[13] ——, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 384–391, Mar. 2003.

[14] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inform. Theory*, pp. 531–532, Jul. 1972.

[15] M. Morelli, A. N. D'Andrea, and U. Mengali, "Feedforward ML-based timing estimation with PSK signals," *IEEE Commun. Lett.*, vol. 1, no. 5, pp. 80–82, May 1997.

[16] M. Oerder and H. Meyr, "Digital filter and square timing recovery," *IEEE Trans. Commun.*, vol. 36, no. 5, pp. 605–611, May 1988.

[17] A. A. D'Amico, A. N. D'Andrea, and U. Mengali, "Feedforward joint phase and timing estimation with OQPSK modulation," *IEEE Trans. Vech. Technol.*, vol. 48, no. 5, pp. 824–832, May 1999.

[18] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.

Yik-Chung Wu received the B.Eng. (honors) and M.Phil. degree in electronic engineering from the University of Hong Kong, Hong Kong, China, in 1998 and 2001, respectively. Currently, he is pursuing the Ph.D. degree at Texas A&M University, College Station.

He was a Research Assistant at the University of Hong Kong, from 2001 to 2002. His research interests include digital signal processing with applications to communication systems, software radio, and space–time processing.

Mr. Wu obtained the Croucher Foundation scholarship in 2002.

S. C. Chan (S'87–M'92) received the B.Sc. (Eng.) and Ph.D. degrees from the University of Hong Kong, Honk Kong, China, in 1986 and 1992, respectively.

He joined City Polytechnic of Hong Kong in 1990 as an Assistant Lecturer and later as a University Lecturer. Since 1994, he has been with the Department of Electrical and Electronic Engineering, University of Hong Kong, and is now an Associate Professor. He was a Visiting Researcher at Microsoft Corporation, USA, and Microsoft, China, in 1998 and 1999, respectively. His research interests include fast transform algorithms, filter design and realization, multirate signal processing, signal processing for communication systems, and image-based rendering.

Dr. Chan is a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society. He was Chairman of the IEEE Hong Kong Chapter of Signal Processing from 2000 to 2002.

Erchin Serpedin (S'95–M'99–SM'04) received (with highest distinction) the Diploma of Electrical Engineering from the Polytechnic Institute of Bucharest, Bucharest, Romania, in 1991, the specialization degree in signal processing and transmission of information from Ecole Superiéure D'Electricité, Paris, France, in 1992, the M.Sc. degree from the Georgia Institute of Technology, Atlanta, in 1992, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, in January 1999.

From 1993 to 1995, he was an Instructor at the Polytechnic Institute of Bucharest, and between January–June 1999, he was a Lecturer at the University of Virginia. In July 1999, he joined the Wireless Communications Laboratory, Texas A&M University, College Station, as an Assistant Professor. His research interests include the areas of statistical signal processing and wireless communications.

Dr. Serpedin has received the NSF Career Award in 2001, and is currently an Associate Editor for the *IEEE Communications Letters*, the *IEEE Signal Processing Letters*, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.