WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2004; **4**:773–790
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/wcm.251

Symbol timing estimation in MIMO correlated flat-fading channels

Yik-Chung Wu and Erchin Serpedin*,†
Department of Electrical Engineering, Texas A&M University, College Station, TX 77843-3128, U.S.A.

Summary

In this paper, the data aided (DA) and non-data aided (NDA) maximum likelihood (ML) symbol timing estimators and their corresponding conditional Cramer–Rao bound (CCRB) and modified Cramer–Rao bound (MCRB) in multiple-input-multiple-output (MIMO) correlated flat-fading channels are derived. It is shown that the approximated ML algorithm in References [4,13] is just a special case of the DA ML estimator; while the extended squaring algorithm in Reference [14] is just a special case of the NDA ML estimator. For the DA case, the optimal orthogonal training sequences are also derived. It is found that the optimal orthogonal sequences resemble the Walsh sequences, but present different envelopes. Simulation results under different operating conditions (e.g. number of antennas and correlation between antennas) are given to assess and compare the performances of the DA and NDA ML estimators with respect to their corresponding CCRBs and MCRBs. It is found that (i) the mean square error (MSE) of the DA ML estimator is close to the CCRB and MCRB, (ii) the MSE of the NDA ML estimator is close to the CCRB but not to the MCRB, (iii) the MSEs of both DA and NDA ML estimators are approximately independent of the number of transmit antennas and are inversely proportional to the number of receive antennas, (iv) correlation between antennas has little effect on the MSEs of DA and NDA ML estimators and (v) DA ML estimator performs better than NDA ML estimator at the cost of lower transmission efficiency and higher implementation complexity. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: maximum likelihood; data-aided; non-data aided; symbol timing estimation; MIMO, correlated fading; Cramer–Rao bound; optimal training sequences

1. Introduction

Communication over multiple-input-multiple-output (MIMO) channels has attracted much attention recently [1–12] due to the huge capacity gain over single antenna systems. While many different techniques and algorithms have been proposed to explore the potential capacity, synchronization in MIMO channels received comparatively less attention.

Symbol timing synchronization in MIMO uncorrelated flat-fading channels was first studied by Naguib *et al.* [4], where the timing delay is estimated by selecting the sample with maximum amplitude from the oversampled approximated log-likelihood

*Correspondence to: Erchin Serpedin, Department of Electrical Engineering, Texas A&M University, College Station, TX 77843-3128, U.S.A.
†E-mail: serpedin@ee.tamu.edu

Contract/grant sponsor: NSF Award; contract/grant number: CCR-0092901.
Contract/grant sponsor: Croucher Foundation.

function. This algorithm was extended by the authors in Reference [13] to increase its estimation accuracy. Unfortunately, the algorithms in References [4,13] are derived in an ad hoc fashion and there is no objective criteria for comparison. On the other hand, the well-known squaring algorithm [24] for symbol timing estimation in single-input-single-output (SISO) channels was extended to MIMO channels in Reference [14], resulting in a non-data aided estimator. However, the estimator proposed in Reference [14] suffers from the problem of self-noise, which is inherited from the original squaring algorithm.

In this paper, the data aided (DA) and non-data aided (NDA) maximum likelihood (ML) symbol timing estimators in MIMO correlated flat-fading channels are derived. In particular, the technique of conditional ML [21,22], in which the nuisance parameters are treated as deterministic but unknown and are estimated together with the parameter of interest, is employed. The advantage of conditional ML method is that there is no need to know or assume the statistical properties of the nuisance parameters. It is shown that the approximated ML algorithm in Reference [4,13] is just a special case of the DA ML estimator; while the extended squaring algorithm in Reference [14] is just a special case of the NDA ML estimator. For the DA case, the optimal orthogonal training sequences are also derived. It is found that the optimal orthogonal training sequences resemble Walsh sequences, but with different envelopes. Two performance bounds are derived for comparison. The first one is the conditional Cramer–Rao bound (CCRB) [18,19], which is the Cramer–Rao bound (CRB) for the symbol timing estimation conditioned that the nuisance parameters are treated as deterministic and are jointly estimated together with the unknown symbol timing. Therefore, the CCRB serves as a performance lower bound for the ML estimators derived. The second one is the modified CRB (MCRB) [20], which is a lower bound for any unbiased symbol timing estimator, irrespective of the underlaying assumption about the nuisance parameters. Being easier to evaluate than CRB, MCRB serves as the ultimate estimation accuracy that may be achieved.

Simulation results under different operating conditions (e.g. number of antennas and correlation between antennas) are given to assess the performances of the DA and NDA ML estimators and compared to the corresponding CCRBs and MCRBs. It is found that (i) the mean square error (MSE) of the DA ML estimator is close to the CCRB and MCRB, meaning that the DA ML estimator is almost the best estimator (in terms of the MSE performance) for the problem under consideration, (ii) the MSE of the NDA ML estimator is close to the CCRB but not MCRB, meaning that NDA ML estimator is an efficient estimator conditioned that the nuisance parameters are being jointly estimated, but there might exist other NDA estimators with better performances; (iii) the MSEs of both DA and NDA ML estimators are approximately independent of the number of transmit antennas and are inversely proportional to the number of receive antennas, (iv) correlation between antennas has little effect on the MSEs of DA and NDA ML estimators unless the correlation coefficient between adjacent antennas is larger than 0.5, in which case small degradation errors occur and v) DA ML estimator performs better than NDA ML estimator at the cost of lower transmission efficiency and higher implementation complexity.

The rest of the paper is organized as follows: the signal model is first described in Section 2. The DA symbol timing estimation problem is addressed in Section 3, in which the ML estimator, the corresponding CCRB and MCRB and the optimal orthogonal training sequences are derived. The NDA ML symbol timing estimator and the corresponding CCRB and MCRB are presented in Section 4. Simulation results are then presented in Section 5 and finally conclusions are drawn in Section 6.

The following notations are used throughout the paper. The symbols $\mathbf{x}^*$, $\mathbf{x}^T$, $\mathbf{x}^H$ and $\|\mathbf{x}\|$ denote the conjugate, transpose, transpose conjugate and the Euclidean norm of $\mathbf{x}$ respectively. Notation $\otimes$ denotes Kronecker products, and vec($\mathbf{H}$) denotes a vector formed by stacking the columns of $\mathbf{H}$, one on top of each other. $\mathbb{E}[\mathbf{x}]$ stands for the expectation of $\mathbf{x}$. Matrices $\mathbf{I}_N$ and $\mathbf{0}_N$ are the identity and the all zero matrix respectively and both are of dimensions $N \times N$. $\mathbf{Z}_{i,:}$, $\mathbf{Z}_{:,j}$ and $\mathbf{Z}_{ij}$ denote the ith row, jth column and $(i,j)^{\text{th}}$ element of $\mathbf{Z}$ respectively. Furthermore, we refer to the DA ML estimator as ML_{DA}, the NDA ML estimator as ML_{NDA} and the corresponding CCRB (MCRB) as CCRB_{DA} (MCRB_{DA}) and CCRB_{NDA} (MCRB_{NDA}) respectively.

2. Signal Model

Consider a MIMO communication system with N transmit and M receive antennas. At each receiving antenna, a superposition of faded signals from all the transmit antennas plus noise is received. Throughout

this paper, it is assumed that the channel is frequency flat and quasi-static. The complex envelope of the received signal at the jth receive antenna can be written as

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

where, E_s/N is the symbol energy; h_{ij} is the complex channel coefficient between the ith transmit antenna and the jth receive antenna; $d_i(n)$ is the zero-mean complex valued symbol transmitted from the ith transmit antenna; $g(t)$ is the transmit filter with unit energy; T is the symbol duration; ε_o is the unknown timing offset, which is assumed to be uniformly distributed in the range $[0, 1)$; and $\eta_j(t)$ is the complex-valued circularly distributed Gaussian white noise at the jth receive antenna, with power density N_o. Notice that the timing offsets between all pairs of transmit and receive antennas are assumed to be the same. This assumption holds when both the transmit and receive antenna array sizes are small.

After passing through the anti-aliasing filter, the received signal is then sampled at rate $f_s = 1/T_s$, where, $T_s \triangleq T/Q$. Note that the oversampling factor Q is determined by the frequency span of $g(t)$; if $g(t)$ is bandlimited to $f = \pm 1/T$ (an example of which is the root raised cosine (RRC) pulse), then $Q = 2$ is sufficient. The received vector $\mathbf{r}_j$, which consists of L_oQ consecutive received samples (L_o is the observation length) from the jth receive antenna, can be expressed as (without loss of generality, we consider the received sequence start at $t = 0$)

$$\mathbf{r}_j = \xi \mathbf{A}_{\varepsilon_o} \mathbf{Z} \mathbf{H}_{j,:}^T + \boldsymbol{\eta}_j \tag{2}$$

where,

$$\xi \triangleq \sqrt{\frac{E_s}{NT}} \tag{3}$$

$$\mathbf{r}_j \triangleq [r_j(0) r_j(T_s) \ldots r_j((L_oQ - 1)T_s)]^T \tag{4}$$

$$\mathbf{A}_{\varepsilon_o} \triangleq [\mathbf{a}_{-L_g}(\varepsilon_o) \mathbf{a}_{-L_g+1}(\varepsilon_o) \ldots \mathbf{a}_{L_o+L_g-1}(\varepsilon_o)] \tag{5}$$

$$\mathbf{a}_i(\varepsilon_o) \triangleq [g(-iT - \varepsilon_o T) g(T_s - iT - \varepsilon_o T) \ldots g((L_oQ - 1)T_s - iT - \varepsilon_o T)]^T \tag{6}$$

$$\mathbf{Z} \triangleq [\mathbf{d}_1 \mathbf{d}_2 \cdots \mathbf{d}_N] \tag{7}$$

$$\mathbf{d}_i \triangleq [d_i(-L_g) d_i(-L_g + 1) \cdots d_i(L_o + L_g - 1)]^T \tag{8}$$

$$\mathbf{H} \triangleq \begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N1} \\ h_{12} & h_{22} & \cdots & h_{N2} \\ \vdots & & & \vdots \\ h_{1M} & h_{2M} & \cdots & h_{NM} \end{bmatrix} \tag{9}$$

$$\boldsymbol{\eta}_j \triangleq [\eta_j(0) \eta_j(1) \ldots \eta_j(L_oQ - 1)]^T \tag{10}$$

with $\eta_j(i) \triangleq \eta_j(iT/Q)$ and L_g denotes the number of symbols affected by the inter-symbol interference (ISI) introduced by one side of $g(t)$. Stacking the received vectors from all the M receive antennas gives

$$\mathbf{r} = \xi (\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o}) \mathrm{vec}(\mathbf{Z}\mathbf{H}^T) + \boldsymbol{\eta} \tag{11}$$

where, $\mathbf{r} \triangleq [\mathbf{r}_1^T \mathbf{r}_2^T \ldots \mathbf{r}_M^T]^T$ and $\boldsymbol{\eta} \triangleq [\boldsymbol{\eta}_1^T \boldsymbol{\eta}_2^T \ldots \boldsymbol{\eta}_M^T]^T$

In order to include the correlation between channel coefficients, the channel transfer function is expressed as:

$$\mathbf{H} = \sqrt{\boldsymbol{\Phi}_R} \mathbf{H}_{\text{i.i.d.}} \sqrt{\boldsymbol{\Phi}_T}^T \tag{12}$$

where, $\boldsymbol{\Phi}_T$ and $\boldsymbol{\Phi}_R$ are the power correlation matrices [10] (normalized such that the diagonal elements are ones) of transmit and receive antenna arrays (which are assumed known) respectively. $\mathbf{H}_{\text{i.i.d.}} \in \mathbb{C}^{M \times N}$ contains independently and identically distributed (i.i.d.) zero-mean, unit-variance, circular symmetric complex Gaussian entries and the matrix square roots denote Cholesky factors such that $\sqrt{\boldsymbol{\Phi}}\sqrt{\boldsymbol{\Phi}}^H = \boldsymbol{\Phi}$. Note that Equation (12) models the correlation among transmit and receive antenna arrays independently. This model is based on the assumption that only immediate surroundings of the antenna array impose the correlation between antenna array elements and have no impact on the correlations at the other end of the communication link. The validity of this model for narrowband nonline-of-sight MIMO channels is verified by recent measurements [7–10]. Substituting Equation (12) into Equation (11), we obtain

$$\mathbf{r} = \xi (\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o}) \mathrm{vec}\left(\mathbf{Z} \sqrt{\boldsymbol{\Phi}_T} \mathbf{H}_{\text{i.i.d.}}^T \sqrt{\boldsymbol{\Phi}_R}^T \right) + \boldsymbol{\eta} \tag{13}$$

3. Symbol Timing Estimation with Known Training Data

3.1. ML Estimator

In this case, the matrix $\mathbf{Z}$ contains the known training sequences and the only unknown is $\mathbf{H}_{\text{i.i.d.}}$. Noting the

fact that $\text{vec}(\mathbf{AYB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}\mathbf{Y}$, then Equation (13) becomes

$$\begin{aligned}\mathbf{r} &= \xi(\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o})(\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{Z}\sqrt{\mathbf{\Phi}_T})\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T) + \boldsymbol{\eta} \\ &= \xi(\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{A}_{\varepsilon_o}\mathbf{Z}\sqrt{\mathbf{\Phi}_T})\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T) + \boldsymbol{\eta}\end{aligned} \tag{14}$$

where, the last line comes from the fact that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$.

From Equation (14), the joint ML estimate of ε_o and $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$ is obtained by maximizing

$$p(\mathbf{r}|\varepsilon, \mathbf{h}) = \frac{1}{(\pi N_o)^{L_o Q}} \exp\left[-\frac{(\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{H})^H (\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{h})}{N_o}\right] \tag{15}$$

or equivalently minimizing

$$J_1(\mathbf{r}|\varepsilon, \mathbf{h}) = (\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{h})^H (\mathbf{r} - \bar{\mathbf{A}}_\varepsilon \mathbf{h}) \tag{16}$$

where, $\bar{\mathbf{A}}_\varepsilon \triangleq \xi(\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{A}_\varepsilon \mathbf{Z}\sqrt{\mathbf{\Phi}_T})$, and ε and $\mathbf{h}$ are the trial values for ε_o and $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$, respectively.

Setting the partial derivatives of $J_1(\mathbf{r}|\varepsilon, \mathbf{h})$ with respect to $\mathbf{h}$ equal to zero, we obtain the ML estimate for $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$ (when ε is fixed) as Reference [15]

$$\hat{\mathbf{h}} = (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H \mathbf{r} \tag{17}$$

Substituting Equation (17) into Equation (16), after some straightforward manipulations and dropping the irrelevant terms, the timing delay is estimated by maximizing the following likelihood function

$$\Lambda_{\text{DA}}(\varepsilon) = \mathbf{r}^H \bar{\mathbf{A}}_\varepsilon (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H \mathbf{r} \tag{18}$$

Using the well-known properties of the Kronecker product $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ and $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$ to expand $\bar{\mathbf{A}}_\varepsilon (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H$, we have

$$\begin{aligned}\bar{\mathbf{A}}_\varepsilon (\bar{\mathbf{A}}_\varepsilon^H \bar{\mathbf{A}}_\varepsilon)^{-1} \bar{\mathbf{A}}_\varepsilon^H &= \left[\sqrt{\mathbf{\Phi}_R}(\sqrt{\mathbf{\Phi}_R}^H \sqrt{\mathbf{\Phi}_R})^{-1} \sqrt{\mathbf{\Phi}_R}^H\right] \\ \otimes \left[\mathbf{A}_\varepsilon \mathbf{Z}\sqrt{\mathbf{\Phi}_T}(\sqrt{\mathbf{\Phi}_T}^H \mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z}\sqrt{\mathbf{\Phi}_T})^{-1} \sqrt{\mathbf{\Phi}_T}^H \mathbf{Z}^H \mathbf{A}_\varepsilon^H\right] \\ &= \mathbf{I}_M \otimes \mathbf{A}_\varepsilon \mathbf{Z}(\mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_\varepsilon^H\end{aligned} \tag{19}$$

where in the second equality, we used the fact that $\sqrt{\mathbf{\Phi}_R}$ and $\sqrt{\mathbf{\Phi}_T}$ are both non-singular square matrices. Substituting this result back into Equation (18), the DA likelihood function is given by

$$\begin{aligned}\Lambda_{\text{DA}}(\varepsilon) &= \mathbf{r}^H (\mathbf{I}_M \otimes \mathbf{A}_\varepsilon \mathbf{Z}(\mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_\varepsilon^H)\mathbf{r} \\ &= \sum_{j=1}^{M} \mathbf{r}_j^H \mathbf{A}_\varepsilon \mathbf{Z}(\mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_\varepsilon^H \mathbf{r}_j\end{aligned} \tag{20}$$

and the ML_{DA} symbol timing estimator can be written as

$$\hat{\varepsilon} = \arg\max_{\varepsilon} \Lambda_{\text{DA}}(\varepsilon) \tag{21}$$

We make the following remarks:

(1) The maximization of the likelihood function usually involves a two-step approach. The first step (coarse search) computes $\Lambda_{\text{DA}}(\varepsilon)$ over a grid of timing delay $\varepsilon_k \triangleq k/K$ for $k = 0, 1, \ldots, K-1$, and then the ε_k that maximizes $\Lambda_{\text{DA}}(\varepsilon)$ is selected. The second step (fine search) finds the global maximum by using either the gradient method [19], dichotomous search [17] or interpolation [17]. In this paper, we employ the parabolic interpolation in the second step due to its implementation simplicity. More specifically, assume that $\Lambda_{\text{DA}}(\varepsilon_{\hat{k}})$ is identified as the maximum among all $\Lambda_{\text{DA}}(\varepsilon_k)$ in the first step. Define $I_1 \triangleq \Lambda_{\text{DA}}(\varepsilon_{\hat{k}-1})$, $I_2 \triangleq \Lambda_{\text{DA}}(\varepsilon_{\hat{k}})$ and $I_3 \triangleq m\Lambda_{\text{DA}}(\varepsilon_{\hat{k}+1})$, then [17]
$$\hat{\varepsilon} = \varepsilon_{\hat{k}} + \frac{I_1 - I_3}{2K(I_1 + I_3 - 2I_2)} \tag{22}$$

(2) The likelihood function at each received antenna can be calculated independently and then added together to obtain the overall likelihood function.

(3) The correlations in the transmit and receive antenna arrays do not appear in the estimator. That is, the ML_{DA} symbol timing estimator is independent of the antenna correlations. This is a reasonable result since another way of deriving the DA likelihood function (20) is not separating $\sqrt{\mathbf{\Phi}_R}$ and $\sqrt{\mathbf{\Phi}_T}$ from $\mathbf{H}_{\text{i.i.d.}}$ and treat $\text{vec}(\mathbf{H}^T)$ as deterministic unknown. Thus, $\mathbf{\Phi}_R$ and $\mathbf{\Phi}_T$ would not appear in the estimator.

(4) In order for the estimate of $\text{vec}(\mathbf{H}_{\text{i.i.d.}}^T)$ to hold in Equation (17), it is required that $\bar{\mathbf{A}}_\varepsilon$ is full rank [15] or equivalently $\sqrt{\mathbf{\Phi}_R}$, $\mathbf{A}_\varepsilon$, $\mathbf{Z}$ and $\sqrt{\mathbf{\Phi}_T}$ are all full rank. Note that $\sqrt{\mathbf{\Phi}_R}$ and $\sqrt{\mathbf{\Phi}_T}$ are lower triangular matrices with positive diagonal elements [16], so they are full rank. Furthermore, if $g(t)$ being a RRC pulse (which is the most

frequently used pulse shape), numerical calculations show that $\mathbf{A}_\varepsilon$ is full rank. Finally, $\mathbf{Z}$ can be made full rank by properly designing the training data. A sufficient condition is that parts of the training sequences from different transmit antennas are orthogonal. That is, for $i \neq j$,

$$[d_i(a)\cdots d_i(b)] \times [d_j(a)\cdots d_j(b)]^H = 0 \qquad (23)$$

for some $a, b \in \{-L_g, -L_g+1, \ldots, L_o+L_g-1\}$ with $a < b$.

(5) For a large observation interval L_o, the (i,j)th element of $\mathbf{A}_\varepsilon^H\mathbf{A}_\varepsilon$ $(i,j = -L_g, L_g+1, \ldots, L_o+L_g-1)$ can be approximated by

$$\begin{aligned}[\mathbf{A}_\varepsilon^H\mathbf{A}_\varepsilon]_{ij} &\approx \sum_{n=-\infty}^{\infty} g^*(nT_s - iT - \varepsilon T) \\ &\times g(nT_s - jT - \varepsilon T) = R_{gg}((i-j)T)\end{aligned} \qquad (24)$$

where, $R_{gg}(\tau)$ is the continuous autocorrelation function of $g(t)$ and the last equality is due to the fact that the sampling rate is at least at the Nyquist rate, which guarantees the equivalence between the discrete and continuous autocorrelation functions of $g(t)$. Therefore, $[\mathbf{A}_\varepsilon^H\mathbf{A}_\varepsilon]_{ij}$ is approximately independent of ε. Note that this approximation is very accurate for the central portion of $\mathbf{A}_\varepsilon^H\mathbf{A}_\varepsilon$. If $R_{gg}(\tau)$ satisfies the Nyquist condition for zero ISI (e.g. $g(t)$ being a RRC pulse or the class of non-bandlimited pulse shapes with $R_{gg}(\tau)$ being time-limited to $[-T/2, T/2]$), then $[\mathbf{A}_\varepsilon^H\mathbf{A}_\varepsilon]_{ij} \approx \delta_{ij}$. Furthermore, if the training sequences from different transmit antennas are orthogonal and with the same norm (i.e. $\mathbf{Z}^H\mathbf{Z} = c\mathbf{I}_N$ for some constant c), then

$$\begin{aligned}\Lambda_{\mathrm{DA}}(\varepsilon) &\approx \frac{1}{c}\sum_{j=1}^{M} \mathbf{r}_j^H\mathbf{A}_\varepsilon\mathbf{Z}\mathbf{Z}^H\mathbf{A}_\varepsilon^H\mathbf{r}_j \\ &= \frac{1}{c}\sum_{j=1}^{M}\sum_{i=1}^{N} |\mathbf{d}_i^H\mathbf{A}_\varepsilon^H\mathbf{r}_j|^2\end{aligned} \qquad (25)$$

Note that $\mathbf{A}_\varepsilon^H\mathbf{r}_j$ is the matched filtering of $\mathbf{r}_j$ with one output sample per symbol with delay ε [21,23]. This reduces to the approximated ML function proposed in Reference [4].

(6) An interesting question is how large L_o is sufficient for the use of Equation (25) in place of Equation (20) without a noticeable loss in performance. The answer depends on the signal-to-noise ratio (SNR) where the estimators work. In general, the higher the SNR, the larger the L_o is required. For example, Figure 1 compares the MSE performances of the true ML estimator and the approximated ML estimator (the training sequences are the optimal orthogonal sequences derived later in this paper). It can be seen that for SNR ≤ 20 dB, $L_o = 32$ is enough for both estimators to have similar performances. For SNR$=$ 30 dB, $L_o = 64$ is required.

(7) In some space-time processing algorithms, (e.g. space-time coding [2–5]), it is required that the channel matrix be also estimated. It is clear that once the timing estimate $\hat{\varepsilon}$ has been found by maximizing Equation (21), the channel estimate can also be obtained readily by using Equation (17). Putting $\varepsilon = \hat{\varepsilon}$ into Equation (17) and expanding it gives

$$\begin{aligned}\hat{\mathbf{h}} = \xi^{-1}\Big((\sqrt{\mathbf{\Phi}_R})^{-1} \otimes (\sqrt{\mathbf{\Phi}_T})^{-1} \\ (\mathbf{Z}^H\mathbf{A}_{\hat{\varepsilon}}^H\mathbf{A}_{\hat{\varepsilon}}\mathbf{Z})^{-1}\mathbf{Z}^H\mathbf{A}_{\hat{\varepsilon}}^H\Big)\mathbf{r}\end{aligned} \qquad (26)$$

If the channel coefficients are uncorrelated (i.e. $\mathbf{\Phi}_T = \mathbf{I}_N$ and $\mathbf{\Phi}_R = \mathbf{I}_M$) and the training sequences from different transmit antennas are orthogonal (i.e. $\mathbf{Z}^H\mathbf{Z} = c\mathbf{I}_N$), it can be easily shown that Equation (26) reduces to

$$\hat{h}_{ij} \approx \frac{1}{c\xi}\mathbf{d}_i^*\mathbf{A}_{\hat{\varepsilon}}^H\mathbf{r}_j \qquad (27)$$

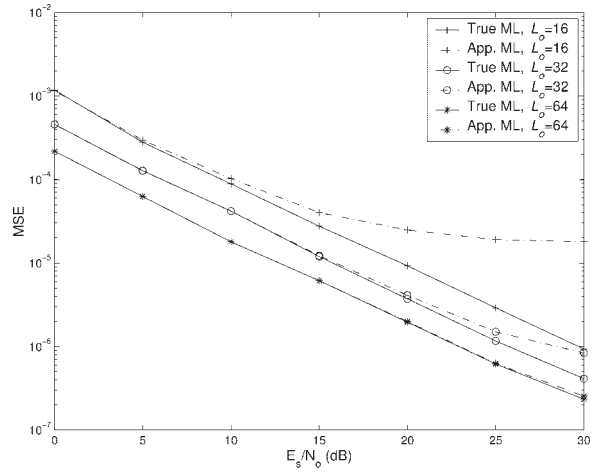

Fig. 1. Mean square error (MSE) performances comparison between the true and approximated data aided maximum likelihood (DA ML) estimators with different $L_o (M = N = 4, L_g = 4, g(t)$ being a root raised cosine (RRC) pulse with roll-off factor $\alpha = 0.3$, $\mathbf{\Phi}_T = \mathbf{I}_4$, $\mathbf{Z} = \mathbf{Z}_{\mathrm{opt}})$.

which is the channel estimation method proposed in Reference [4].

3.2. The CCRB and MCRB

For the model in Equation (14), it is known that for a specific timing delay ε_o, the $\mathrm{CCRB}_{\mathrm{DA}}$ is given by[‡] Reference [22].

$$\mathrm{CCRB}_{\mathrm{DA}}(\varepsilon_o) = \frac{\sigma^2}{2\mathrm{tr}(\bar{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_{\bar{\mathbf{A}}}^{\perp} \bar{\mathbf{D}}_{\varepsilon_o} \boldsymbol{\Gamma}_{\mathbf{h}})} \tag{28}$$

In Equation (28), $\sigma^2 = N_o f_s = N_o Q/T$ is the noise variance, tr(.) denotes the trace of a matrix,

$$\bar{\mathbf{D}}_{\varepsilon} \triangleq \frac{d\bar{\mathbf{A}}_{\varepsilon}}{d\varepsilon} = \xi \sqrt{\boldsymbol{\Phi}_R} \otimes \mathbf{D}_{\varepsilon} \mathbf{Z} \sqrt{\boldsymbol{\Phi}_T} \tag{29}$$

with $\mathbf{D}_{\varepsilon} \triangleq \mathrm{d}\mathbf{A}_{\varepsilon}/\mathrm{d}\varepsilon$, $\mathbf{P}_{\bar{\mathbf{A}}}^{\perp}$ is the orthogonal projector onto the null space of $\bar{\mathbf{A}}_{\varepsilon_o}$ and is given by

$$\begin{aligned} \mathbf{P}_{\bar{\mathbf{A}}}^{\perp} &\triangleq \mathbf{I}_{ML_oQ} - \bar{\mathbf{A}}_{\varepsilon_o} (\bar{\mathbf{A}}_{\varepsilon_o}^H \bar{\mathbf{A}}_{\varepsilon_o})^{-1} \bar{\mathbf{A}}_{\varepsilon_o}^H \\ &= \mathbf{I}_M \otimes (\mathbf{I}_{L_oQ} - \mathbf{A}_{\varepsilon_o} \mathbf{Z} (\mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H \mathbf{A}_{\varepsilon_o} \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H) \\ &= \mathbf{I}_M \otimes \mathbf{P}_{\mathbf{AZ}}^{\perp} \end{aligned} \tag{30}$$

where, $\mathbf{P}_{\mathbf{AZ}}^{\perp} \triangleq \mathbf{I}_{L_oQ} - \mathbf{A}_{\varepsilon_o} \mathbf{Z} (\mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H \mathbf{A}_{\varepsilon_o} \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{A}_{\varepsilon_o}^H$, and

$$\boldsymbol{\Gamma}_{\mathbf{h}} \triangleq \mathbb{E}[\mathrm{vec}(\mathbf{H}_{\mathrm{i.i.d.}}^T)\mathrm{vec}(\mathbf{H}_{\mathrm{i.i.d.}}^T)^H] = \mathbf{I}_{MN} = \mathbf{I}_M \otimes \mathbf{I}_N \tag{31}$$

Subsituting Equations (29), (30) and (31) into Equation (28), we obtain

$$\begin{aligned} \mathrm{CCRB}_{\mathrm{DA}}(\varepsilon_o) &= \frac{\sigma^2}{2\xi^2 \mathrm{tr}\left((\sqrt{\boldsymbol{\Phi}_R} \otimes \mathbf{D}_{\varepsilon_o} \mathbf{Z} \sqrt{\boldsymbol{\Phi}_T})^H (\mathbf{I}_M \otimes \mathbf{P}_{\mathbf{AZ}}^{\perp}) (\sqrt{\boldsymbol{\Phi}_R} \otimes \mathbf{D}_{\varepsilon_o} \mathbf{Z} \sqrt{\boldsymbol{\Phi}_T}) (\mathbf{I}_M \otimes \mathbf{I}_N) \right)} \\ &= \frac{QN}{2\mathrm{tr}(\sqrt{\boldsymbol{\Phi}_R}^H \sqrt{\boldsymbol{\Phi}_R}) \mathrm{tr}(\sqrt{\boldsymbol{\Phi}_T}^H \mathbf{Z}^H \mathbf{D}_{\varepsilon_o}^H \mathbf{P}_{\mathbf{AZ}}^{\perp} \mathbf{D}_{\varepsilon_o} \mathbf{Z} \sqrt{\boldsymbol{\Phi}_T})} \left(\frac{E_s}{N_o} \right)^{-1} \\ &= \frac{1}{2M \mathrm{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_{\mathbf{AZ}}^{\perp} \tilde{\mathbf{D}}_{\varepsilon_o} \tilde{\mathbf{Z}} \boldsymbol{\Phi}_T)} \left(\frac{E_s}{N_o} \right)^{-1} \end{aligned} \tag{32}$$

where, $\tilde{\mathbf{Z}} \triangleq \mathbf{Z}/\sqrt{N}$ and $\tilde{\mathbf{D}}_{\varepsilon} \triangleq \mathbf{D}_{\varepsilon}/\sqrt{Q}$. In passing from the second line to the third line in Equation (32), we used the fact that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ and the diagonal elements of $\boldsymbol{\Phi}_R$ are all one regardless of the specific value of the correlation matrix.

For a specific timing delay ε_o, $\mathrm{MCRB}_{\mathrm{DA}}$ is given by Reference [22]

$$\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o) = \frac{\sigma^2}{2\mathrm{tr}(\bar{\mathbf{D}}_{\varepsilon_o}^H \bar{\mathbf{D}}_{\varepsilon_o} \boldsymbol{\Gamma}_{\mathbf{h}})} \tag{33}$$

and based on similar calculations with those used for $\mathrm{CCRB}_{\mathrm{DA}}$, it can be shown that

$$\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o) = \frac{1}{2M \mathrm{tr}(\tilde{\mathbf{Z}}^H \tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o} \tilde{\mathbf{Z}} \boldsymbol{\Phi}_T)} \left(\frac{E_s}{N_o} \right)^{-1} \tag{34}$$

The following remarks concerning the $\mathrm{CCRB}_{\mathrm{DA}}$ and $\mathrm{MCRB}_{\mathrm{DA}}$ are now in order:

(1) Since the timing delay ε_o is assumed uniformly distributed, the average of $\mathrm{CCRB}_{\mathrm{DA}}$ and $\mathrm{MCRB}_{\mathrm{DA}}$ can be calculated by numerical integration of Equations (32) and (34) respectively.
(2) The $\mathrm{CCRB}_{\mathrm{DA}}$ and $\mathrm{MCRB}_{\mathrm{DA}}$ do not depend on the receive antenna array correlation matrix $\boldsymbol{\Phi}_R$. Furthermore, the $\mathrm{CCRB}_{\mathrm{DA}}$ and $\mathrm{MCRB}_{\mathrm{DA}}$ are inversely proportional to the number of receive antennas M. Thus, the $\mathrm{CCRB}_{\mathrm{DA}}$ and $\mathrm{MCRB}_{\mathrm{DA}}$ will be reduced by a factor of 2 whenever the number of receive antennas M is doubled.
(3) The expressions for $\mathrm{CCRB}_{\mathrm{DA}}$ and $\mathrm{MCRB}_{\mathrm{DA}}$ would still be given by Equations (32) and (34) respectively, even if we treat $\mathrm{vec}(\mathbf{H}^T)$ as deterministic unknown rather than $\mathrm{vec}(\mathbf{H}_{\mathrm{i.i.d.}}^T)$ in the system model.

3.3. Optimal Orthogonal Training Sequences

Since the $\mathrm{CCRB}_{\mathrm{DA}}$ can be reached asymptotically by the $\mathrm{ML}_{\mathrm{DA}}$ estimator, (21) [15], it is natural to search for optimal training sequences by minimizing the

[‡]Strictly speaking, the bound given is the asymptotic CCRB. However, it is shown in Reference [22] that the true CCRB tends to the asymptotic CCRB, when $M, N \to \infty$.

$\mathrm{CCRB}_{\mathrm{DA}}$ in Equation (32) with respect to $\mathbf{Z}$. Unfortunately, since the denominator of Equation (32) is a very complicated function of $\mathbf{Z}$, it is difficult, if not impossible, to obtain a simple solution. On the other hand, the expression for the $\mathrm{MCRB}_{\mathrm{DA}}$ in Equation (34) has a much simpler dependence on $\mathbf{Z}$. Moreover, it will be shown later in this section that for the derived optimal training sequences, the corresponding $\mathrm{CCRB}_{\mathrm{DA}}$ is actually very close to that of $\mathrm{MCRB}_{\mathrm{DA}}$ (see Figure 3). Therefore, in the following the optimal training sequences are derived by minimizing the $\mathrm{MCRB}_{\mathrm{DA}}$ with respect to $\mathbf{Z}$.

With the constraint that the columns of $\mathbf{Z}$ has to be orthogonal[§] (i.e. $\mathbf{Z}^H\mathbf{Z} = (L_o + 2L_g)\mathbf{I}_N$), it is proved in Appendix I that the matrix $\mathbf{Z}$ that minimizes $\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o)$ is given by

$$\mathbf{Z} = \sqrt{(L_o + 2L_g)}\tilde{\mathbf{U}}(\varepsilon_o)\mathbf{U}_T^H \tag{35}$$

where, $\tilde{U}(\varepsilon_o)$ is the matrix containing the N eigenvectors corresponding to the N largest eigenvalues of $\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}$ as columns and $\mathbf{U}_T$ is the unitary matrix containing all the eigenvectors of $\mathbf{\Phi}_T$ as columns.

In general, the optimal orthogonal training sequences depend on the unknown parameter ε_o. However, note that, following the same argument as in Equation (24), $[\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}]_{ij} \approx R_{g'g'}((i-j)T)T^2/Q$, where, $g'(t) = \mathrm{d}g(t)/\mathrm{d}t$. Therefore, $\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}$ is approximately independent of the parameter ε_o and in practice, we can fix a nominal timing delay, say $\varepsilon_t = 0$ (actually other values do not make a large difference as we will show), for designing the training sequences. This idea is verified by Figure 2, where,

$$\beta \triangleq \frac{1}{\mathrm{tr}(\tilde{\mathbf{Z}}^H\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}\tilde{\mathbf{Z}}\mathbf{\Phi}_T)} \tag{36}$$

is plotted against ε_o for $\varepsilon_t = 0, 0.25, 0.5, 0.75$ with $N = 4, L_o = 32, L_g = 4$, $g(t)$ being a RRC pulse with roll-off factor $\alpha = 0.3$ and $\mathbf{\Phi}_T = \mathbf{I}_4$. The case of $\varepsilon_t = \varepsilon_o$ is also shown for a reference. It is obvious that the mismatch of ε_t and ε_o does not increase the value of β significantly. From Figure 2, we note that the worst case increase of β due to the mismatch of ε_t and ε_o is about 2×10^{-5} and when $\varepsilon_t = \varepsilon_o$,

[§]Notice that in this paper, the search for optimal training sequences would be confined to the class of orthogonal sequences. The question of whether there exists any non-orthogonal training sequences with better performances and how to find them is a subject open to future investigations.

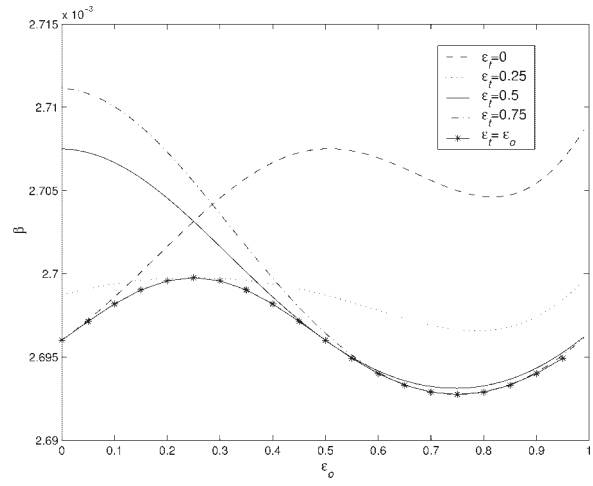

Fig. 2. Plots of $\beta \triangleq 1/\mathrm{tr}(\tilde{\mathbf{Z}}^H\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}\tilde{\mathbf{Z}}\mathbf{\Phi}_T)$ against ε_o for $\varepsilon_t = 0, 0.25, 0.5, 0.75$ ($N = 4, L_o = 32, L_g = 4$, $g(t)$ being a RRC pulse with $\alpha = 0.3$, $\mathbf{\Phi}_T = \mathbf{I}_4$).

$\beta \approx 2.695 \times 10^{-3}$. Thus, the worst case relative error for the $\mathrm{MCRB}_{\mathrm{DA}}$ in this example is

$$\frac{\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o|\varepsilon_t \neq \varepsilon_o) - \mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o|\varepsilon_t = \varepsilon_o)}{\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o|\varepsilon_t = \varepsilon_o)} \approx \frac{2 \times 10^{-5}}{2.695 \times 10^{-3}} = 7.42 \times 10^{-3} \tag{37}$$

The implication of the above calculation is that the worst case variation of the $\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o)$ due to the mismatch between ε_o and ε_t is at least 100 times smaller than the value of the $\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o)$ when $\varepsilon_t = \varepsilon_o$. Therefore, the optimality of the orthogonal training sequences derived is approximately independent of ε_o and we can write $\mathbf{Z}_{\mathrm{opt}} = \sqrt{(L_o + 2L_g)}\tilde{\mathbf{U}}(0)\mathbf{U}_T^H$.

With the optimal orthogonal training sequences $\mathbf{Z}_{\mathrm{opt}}$, the ratios $\mathrm{CCRB}_{\mathrm{DA}}(\varepsilon_o)/\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o)$ are plotted in Figure 3 against the number of transmit antenna N for $\varepsilon_o = 0, 0.25, 0.5, 0.75$ with $L_o = 32$ and $128, L_g = 4$, $g(t)$ being a RRC pulse with $\alpha = 0.3$ and $\mathbf{\Phi}_T = \mathbf{I}_N$. It can be seen that the ratios $\mathrm{CCRB}_{\mathrm{DA}}(\varepsilon_o)/\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o)$ for different ε_o are close to 1 (this is true for the case $L_o = 128$, and for moderate number of transmit antennas when $L_o = 32$). Since, $\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o) \leq \mathrm{CCRB}_{\mathrm{DA}}(\varepsilon_o)$, even there are some other orthogonal sequences that actually minimize the $\mathrm{CCRB}_{\mathrm{DA}}(\varepsilon_o)$, the space for performance improvement is very small (e.g. for $L_o = 32$ and $N \leq 4$, the ratio $\mathrm{CCRB}_{\mathrm{DA}}(\varepsilon_o)/\mathrm{MCRB}_{\mathrm{DA}}(\varepsilon_o)$ is

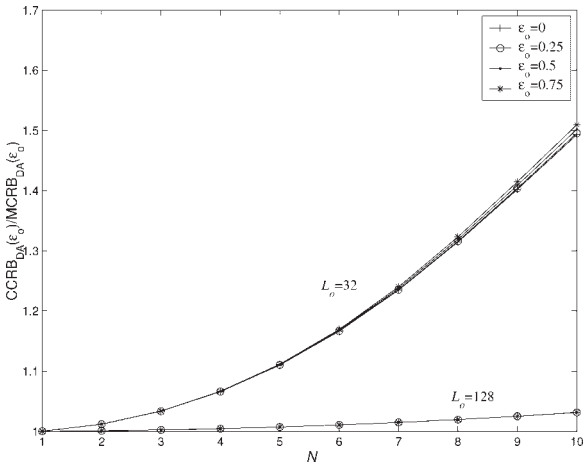

Fig. 3. Plots of conditional Cramer–Rao bound $(\mathrm{CCRB}_{\mathrm{DA}})(\varepsilon_o)$/ modified Cramer–Rao bound $(\mathrm{MCRB}_{\mathrm{DA}})(\varepsilon_o)$ against the number of transmit antenna N for $\varepsilon_o = 0, 0.25, 0.5, 0.75$ ($L_o = 32$ and 128, $L_g = 4$ and $g(t)$ being a RRC pulse with $\alpha = 0.3$, $\mathbf{\Phi}_T = \mathbf{I}_N$, $\mathbf{Z} = \mathbf{Z}_{\mathrm{opt}}$).

smaller than 1.1, the best possible performance improvement is only $10\log_{10}(1.1) \approx 0.4$dB), not mentioning that these training sequences are difficult to find or may even do not exist. This justifies the search for optimal orthogonal training sequences by minimizing the $\mathrm{MCRB}_{\mathrm{DA}}$.

It is interesting to find that, when $\mathbf{\Phi}_T = \mathbf{I}_N$ and $g(t)$ is a RRC pulse, the optimal orthogonal training sequences resemble the Walsh sequences. Let, $\mathbf{w}_n$ be the Walsh sequence with length 32 and with n sign changes. For comparison, Figures 4 and 5 show $[\mathbf{Z}_{\mathrm{opt}}]_{:,1}$ and $[\mathbf{Z}_{\mathrm{opt}}]_{:,2}$ with $L_o = 32, L_g = 4$ and $\alpha = 0.3$, together with $\mathbf{w}_{31}$ and $-\mathbf{w}_{30}$ plotted from the index 5 to 36. Note that the lines are drawn for

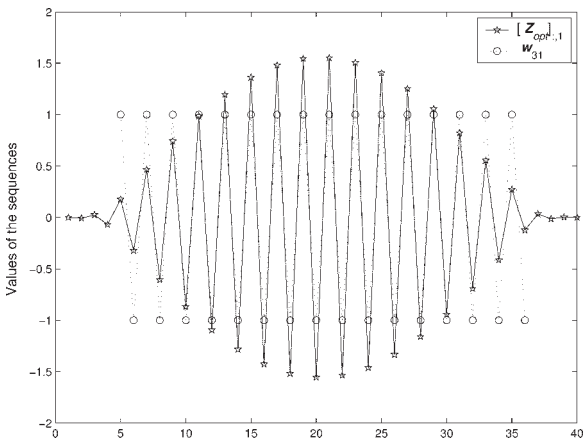

Fig. 4. Plots of $[\mathbf{Z}_{\mathrm{opt}}]_{:,1}$ and $\mathbf{w}_{31}$ ($g(t)$ being a RRC pulse with $\alpha = 0.3$, $L_o = 32$, $L_g = 4$).

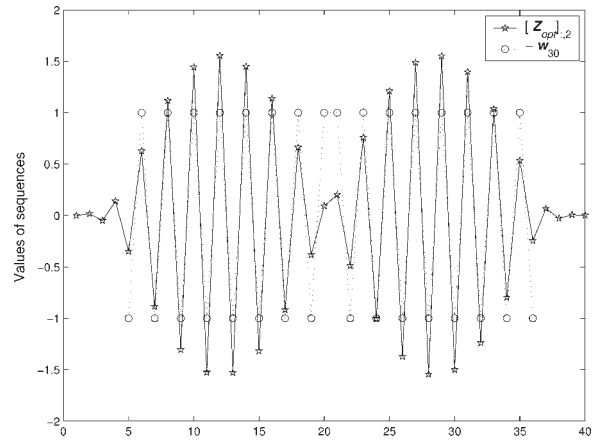

Fig. 5. Plots of $[\mathbf{Z}_{\mathrm{opt}}]_{:,2}$ and $-\mathbf{w}_{30}$ ($g(t)$ being a RRC pulse with $\alpha = 0.3$, $L_o = 32, L_g = 4$).

easy reading, there is no value defined in between integer indexes. It can be observed that, the values of the optimal sequences at indices 1–4 and 37–40 are very small. Moreover, with the exception of the different envelope shapings, the sign-changing patterns of the optimal orthogonal sequences follow that of Walsh sequences (for indices 5–36). In general, the same relationship can be found between $[\mathbf{Z}_{\mathrm{opt}}]_{:,i}$ and $\mathbf{w}_{32-i}$. We also remark that the use of Walsh sequences with the largest number of sign changes for symbol timing estimation in space-time coding system has been initially proposed in Reference [14].

Finally, Figure 6 compares the performance of $\mathrm{ML}_{\mathrm{DA}}$, Equation (21) with different kinds of training sequences in a 4-transmit, 4-receive antenna system

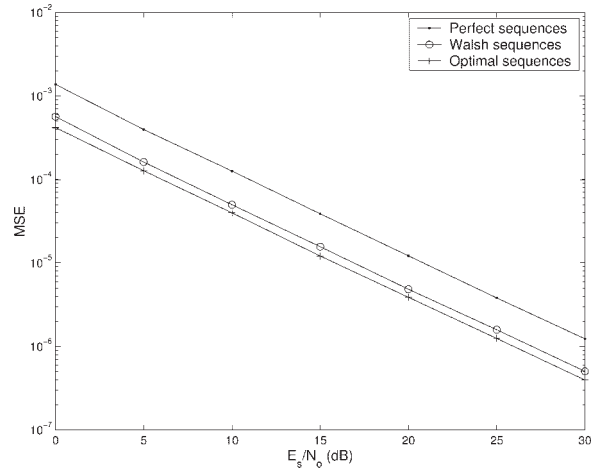

Fig. 6. Comparison of the MSE performances of $\mathrm{ML}_{\mathrm{DA}}$ with different training sequences ($g(t)$ being a RRC pulse with $\alpha = 0.3$, $M = N = 4$, $L_o = 32$, $L_g = 4$, $\mathbf{\Phi}_T = \mathbf{\Phi}_R = \mathbf{I}_4$).

with $L_o = 32, L_g = 4$, $g(t)$ being a RRC pulse with $\alpha = 0.3$. For simplicity, we set $\mathbf{\Phi}_T = \mathbf{\Phi}_R = \mathbf{I}_4$. Three different kinds of training sequences are considered. The first one is the optimal orthogonal training sequences derived above. The second one is the Walsh sequences $\mathbf{w}_{31}, \mathbf{w}_{30}, \mathbf{w}_{29}, \mathbf{w}_{28}$ and extended to length 40 by adding a cyclic prefix and suffix, each of length equal to 4. The final one is the perfect sequences proposed in Reference [13], where they were derived to minimize the contribution of the ISI term in the approximated log-likelihood function (25) (see Reference [13] for detail). From Figure 6, it can be seen that the perfect sequences perform not as well as the Walsh sequences and the optimal sequences. This is because the true ML estimator is used in simulations and the perfect sequences (which were derived based on the approximated log-likelihood function) may not have any optimality. Due to the resemblance of the optimal orthogonal sequences and the Walsh sequences, the performance of the $\mathrm{ML_{DA}}$ by using these two kinds of sequences are close to each other, with the case of optimal orthogonal sequences performing marginally better. For fair comparison, we mention that the perfect sequences and the Walsh sequences are constant modulus sequences while the optimal orthogonal sequences are not.

4. Non-Data Aided Symbol Timing Estimation

4.1. ML Estimator

In this case, no training sequence is used and $\mathbf{Z}$ contains real data. Now, the matrices $\mathbf{Z}$ and $\mathbf{H}_{\text{i.i.d.}}$ in Equation (13) are unknown and Equation (13) can be rewritten in the following form:

$$\begin{aligned}\mathbf{r} &= \xi(\mathbf{I}_M \otimes \mathbf{A}_{\varepsilon_o})(\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{I}_{L_o+2L_g})\mathrm{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T) + \boldsymbol{\eta} \\ &= \xi(\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{A}_{\varepsilon_o})\mathrm{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T) + \boldsymbol{\eta}\end{aligned} \tag{38}$$

Note that although, $\mathbf{\Phi}_T$ is assumed to be known, it cannot be separated from $\mathbf{Z}$ and $\mathbf{H}_{\text{i.i.d.}}$ because the correlation in transmit antennas can be translated into correlation of unknown data or vice versa. Since the noise is white and Gaussian, the $\mathrm{ML_{NDA}}$ estimator resumes to the minimization of

$$J_2(\mathbf{r}|\varepsilon, \mathbf{x}) = (\mathbf{r} - \check{\mathbf{A}}_\varepsilon \mathbf{x})^H (\mathbf{r} - \check{\mathbf{A}}_\varepsilon \mathbf{x}) \tag{39}$$

where, $\check{\mathbf{A}}_\varepsilon \triangleq \xi(\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{A}_\varepsilon)$, ε and $\mathbf{x}$ are the trial values for ε_o and $\mathrm{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T)$ respectively.

With the linear model of Equation (38), the ML estimate for $\mathrm{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T)$ (when ε is fixed) is given by

$$\hat{\mathbf{x}} = (\check{\mathbf{A}}_\varepsilon^H \check{\mathbf{A}}_\varepsilon)^{-1} \check{\mathbf{A}}_\varepsilon^H \mathbf{r} \tag{40}$$

Putting Equation (40) into Equation (39), after some straightforward calculations and dropping the irrelevant terms, the $\mathrm{ML_{NDA}}$ symbol timing estimator reduces to the maximization of the following likelihood function:

$$\Lambda_{\mathrm{NDA}}(\varepsilon) = \mathbf{r}^H \check{\mathbf{A}}_\varepsilon (\check{\mathbf{A}}_\varepsilon^H \check{\mathbf{A}}_\varepsilon)^{-1} \check{\mathbf{A}}_\varepsilon^H \mathbf{r} \tag{41}$$

It can be easily shown that

$$\check{\mathbf{A}}_\varepsilon (\check{\mathbf{A}}_\varepsilon^H \check{\mathbf{A}}_\varepsilon)^{-1} \check{\mathbf{A}}_\varepsilon^H = \mathbf{I}_M \otimes \mathbf{A}_\varepsilon (\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H \tag{42}$$

which gives

$$\Lambda_{\mathrm{NDA}}(\varepsilon) = \sum_{j=1}^{M} \mathbf{r}_j^H \mathbf{A}_\varepsilon (\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon)^{-1} \mathbf{A}_\varepsilon^H \mathbf{r}_j \tag{43}$$

The $\mathrm{ML_{NDA}}$ symbol timing estimation can be stated as

$$\hat{\varepsilon} = \arg\max_{\varepsilon} \Lambda_{\mathrm{NDA}}(\varepsilon) \tag{44}$$

and can be implemented by the two-step approach as for the $\mathrm{ML_{DA}}$.

Note that the implementation of the $\mathrm{ML_{NDA}}$ estimator does not require the knowledge of correlation among antennas. Note also that the likelihood function in Equation (43) is the sum of individual likelihood functions for each receive antenna, just as the case of training-based likelihood function in Equation (20). For each of the receive antenna, the likelihood function is the same as the likelihood function for SISO systems [22,23]. Furthermore, applying the low-complexity maximization technique [23] to the likelihood function (43) and with the approximation $\mathbf{A}_\varepsilon^H \mathbf{A}_\varepsilon \approx \mathbf{I}_{L_o+2L_g}$ for Nyquist zero-ISI pulse, it can be easily shown that the $\mathrm{ML_{NDA}}$ (44) reduces to the extension of squaring algorithm proposed in Reference [14].

4.2. The CCRB and MCRB

For the model in Equation (38), the CCRB for a specific ε_o is given by Reference [21]

$$\mathrm{CCRB_{NDA}}(\varepsilon_o) = \frac{\sigma^2}{2\mathrm{tr}(\check{\mathbf{D}}_{\varepsilon_o}^H \mathbf{P}_{\check{\mathbf{A}}}^{\perp} \check{\mathbf{D}}_{\varepsilon_o} \mathbf{\Gamma}_{\mathbf{x}})} \tag{45}$$

where,

$$\check{\mathbf{D}}_{\varepsilon} \triangleq \frac{\mathrm{d}\check{\mathbf{A}}_{\varepsilon}}{\mathrm{d}\varepsilon} = \sqrt{\mathbf{\Phi}_R} \otimes \mathbf{D}_{\varepsilon} \tag{46}$$

$$\mathbf{P}_{\check{\mathbf{A}}}^{\perp} \triangleq \mathbf{I}_{ML_oQ} - \check{\mathbf{A}}_{\varepsilon_o}(\check{\mathbf{A}}_{\varepsilon_o}{}^{H}\check{\mathbf{A}}_{\varepsilon_o})^{-1}\check{\mathbf{A}}_{\varepsilon_o}{}^{H} = \mathbf{I}_M \otimes \mathbf{P}_{\mathbf{A}}^{\perp} \tag{47}$$

with $\mathbf{P}_{\mathbf{A}}^{\perp} \triangleq \mathbf{I}_{L_oQ} - \mathbf{A}_{\varepsilon_o}(\mathbf{A}_{\varepsilon_o}^{H}\mathbf{A}_{\varepsilon_o})^{-1}\mathbf{A}_{\varepsilon_o}^{H}$, and

$$\mathbf{\Gamma}_{\mathbf{x}} \triangleq \mathbb{E}[\mathrm{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^{T})\,\mathrm{vec}(\mathbf{Z}\sqrt{\mathbf{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^{T})^{H}] \tag{48}$$

It is shown in Appendix II that

$$\mathbf{\Gamma}_{\mathbf{x}} = \mathbf{I}_M \otimes \mathbf{\Psi} \tag{49}$$

where $\mathbf{\Psi}$ is a Hermitian and Toeplitz matrix with elements $[\mathbf{\Psi}]_{ij} \triangleq \mathrm{tr}(\mathbf{\Gamma}_z(j-i)\mathbf{\Phi}_T)$ and $\mathbf{\Gamma}_z(j-i) \triangleq \mathbb{E}[(\mathbf{Z}^*)_{j,:}^{T}(\mathbf{Z})_{i,:}]$ is the average cross-correlation matrix of the symbols transmitted with time index difference $j-i$.

Substituting Equations (46), (47) and (49) into Equation (45), we obtain

$$\mathrm{CCRB}_{\mathrm{NDA}}(\varepsilon_o) = \frac{\sigma^2}{2\xi^2\mathrm{tr}\big((\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{D}_{\varepsilon_o})^{H}(\mathbf{I}_M \otimes \mathbf{P}_{\mathbf{A}}^{\perp})(\sqrt{\mathbf{\Phi}_R} \otimes \mathbf{D}_{\varepsilon_o})(\mathbf{I}_M \otimes \mathbf{\Psi})} \tag{50}$$

$$= \frac{1}{2M\mathrm{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^{H}\mathbf{P}_{\mathbf{A}}^{\perp}\tilde{\mathbf{D}}_{\varepsilon_o}\mathbf{\Psi}/N)}\left(\frac{E_s}{N_o}\right)^{-1} \tag{51}$$

Following the same calculations as for the $\mathrm{CCRB}_{\mathrm{NDA}}$, the $\mathrm{MCRB}_{\mathrm{NDA}}$ is given by

$$\mathrm{MCRB}_{\mathrm{NDA}}(\varepsilon_o) = \frac{\sigma^2}{2\mathrm{tr}(\check{\mathbf{D}}_{\varepsilon_o}^{H}\check{\mathbf{D}}_{\varepsilon_o}\mathbf{\Gamma}_{\mathbf{x}})} \tag{52}$$

$$= \frac{1}{2M\mathrm{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^{H}\tilde{\mathbf{D}}_{\varepsilon_o}\mathbf{\Psi}/N)}\left(\frac{E_s}{N_o}\right)^{-1} \tag{53}$$

Note that the average of $\mathrm{CCRB}_{\mathrm{NDA}}$ and $\mathrm{MCRB}_{\mathrm{NDA}}$ can be computed by numerical integration of Equations (51) and (53) respectively. In the following, we consider two special cases.

Special Case 1: The data is spatially and temporally white (e.g. Vertical-Bell Labs Layered Space-Time (V-BLAST) system[¶] [12]). In this case, $\mathbf{\Gamma}_z(j-i) = \mathbf{I}_N\delta_{ij}$, implying that $[\mathbf{\Psi}]_{ij} = \delta_{ij}\mathrm{tr}(\mathbf{\Phi}_T) = N\delta_{ij}$. Therefore, the corresponding $\mathrm{CCRB}_{\mathrm{NDA}}$ and $\mathrm{MCRB}_{\mathrm{NDA}}$ are

$$\mathrm{CCRB}_{\mathrm{NDA}}(\varepsilon_o) = \frac{1}{2M\mathrm{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^{H}\mathbf{P}_{\mathbf{A}}^{\perp}\tilde{\mathbf{D}}_{\varepsilon_o})}\left(\frac{E_s}{N_o}\right)^{-1} \tag{54}$$

and

$$\mathrm{MCRB}_{\mathrm{NDA}}(\varepsilon_o) = \frac{1}{2M\mathrm{tr}(\tilde{\mathbf{D}}_{\varepsilon_o}^{H}\tilde{\mathbf{D}}_{\varepsilon_o})}\left(\frac{E_s}{N_o}\right)^{-1} \tag{55}$$

respectively. Note that in this case, the $\mathrm{CCRB}_{\mathrm{NDA}}$ and $\mathrm{MCRB}_{\mathrm{NDA}}$ do not depend on the number of transmit antennas and the correlations among antennas.

Special Case 2: Space-time block code (STBC) system. In general, a block of STBC symbols can be represented by a $s \times N$ matrix [6]

$$\mathcal{G} = \sum_{k=1}^{rs}\Re(b_k)\mathbf{X}_k + j\sum_{k=1}^{rs}\Im(b_k)\mathbf{Y}_k \tag{56}$$

where, r is the rate of the STBC, s is the length of the STBC, b_k's are the i.i.d., complex valued symbols to be encoded, $\Re(.)$ and $\Im(.)$ denote the real and imaginary parts, $j \triangleq \sqrt{-1}$ and $\mathbf{X}_k, \mathbf{Y}_k$ are the fixed, real-valued elementary code matrices. Without loss of

[¶]In its initial development, V-BLAST system does not employ any temporal error control code. Although, temporal error control code may be applied in V-BLAST system, we assume the data is temporally white since from the point of view of the symbol synchronizer, the data appears to be uncorrelated.

generality, we assume $|b_k| = 1$. It is proved in Appendix III that for the STBC system,

$$\mathbf{\Gamma}_z(j-i) = \begin{cases} \mathbf{0}_N & \text{for} \quad |j-i| \geq s \\ \frac{1}{2s}\sum_{n=1}^{s-\ell}\left(\sum_{k=1}^{rs}[\mathbf{X}_k]_{n-\ell,:}^T[\mathbf{X}_k]_{n,:} + \sum_{k=1}^{rs}[\mathbf{Y}_k]_{n+\ell,:}^T[\mathbf{Y}_k]_{n,:}\right) & \text{for} \quad |j-i| = \ell, \ell < s \end{cases} \tag{57}$$

For example, let us consider the half-rate orthogonal STBC with four transmit antennas [5], in which case $N = 4, s = 8, r = 1/2$ and the matrix $\mathcal{G}$ given by

$$\mathcal{G} = \begin{pmatrix} b_1 & b_2 & b_3 & b_4 \\ -b_2 & b_1 & -b_4 & b_3 \\ -b_3 & b_4 & b_1 & -b_2 \\ -b_4 & -b_3 & b_2 & b_1 \\ b_1^* & b_2^* & b_3^* & b_4^* \\ -b_2^* & b_1^* & -b_4^* & b_3^* \\ -b_3^* & b_4^* & b_1^* & -b_2^* \\ -b_4^* & -b_3^* & b_2^* & b_1^* \end{pmatrix} \tag{58}$$

Decomposing $\mathcal{G}$ in terms of $\mathbf{X}_k$ and $\mathbf{Y}_k$ and using Equation (57), it is found that

$$\mathbf{\Gamma}_z(j-i) = \begin{cases} \mathbf{I}_4 & \text{for} \quad i = j \\ \frac{1}{4}\begin{bmatrix} 0 & 2 & 0 & 1 \\ -2 & 0 & 1 & 0 \\ 0 & -1 & 0 & 2 \\ -1 & 0 & -2 & 0 \end{bmatrix} & \text{for} \quad |j-i| = 1 \\ \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} & \text{for} \quad |j-i| = 3 \\ \mathbf{0}_4 & \text{otherwise} \end{cases} \tag{59}$$

Then, $\mathbf{\Psi}$ can be computed according to $[\mathbf{\Psi}]_{ij} = \text{tr}(\mathbf{\Gamma}_z(j-i)\mathbf{\Phi}_T)$ and the CCRB_{NDA} and MCRB_{NDA} are given by Equations (51) and (53) respectively.

5. Simulation Results and Discussions

In this section, the MSE performances of the proposed symbol timing estimators, ML_{DA} (21) and ML_{NDA} (44) are assessed by Monte Carlo simulations. In all the simulations, $L_o = 32$, $L_g = 4$ (i.e. the total length of training data is 40), $Q = 2$, $K = 16$, ε_o is uniformly distributed in the range $[0, 1)$ and $g(t)$ is a RRC filter with roll-off factor $\alpha = 0.3$. Each point is obtained by averaging 10^4 Monte-Carlo simulation runs. For the DA case, the optimal orthogonal sequences $\mathbf{Z}_{\text{opt}}$ derived in Section 3.3 are used as training data. For the NDA case, the data format is QPSK.

5.1. Effects of N and M

In this Section, the effects of the number of transmit and receive antennas are examined. First, let us assume, $\mathbf{\Phi}_T = \mathbf{I}_N$ and $\mathbf{\Phi}_R = \mathbf{I}_M$ for the moment. Furthermore, it is assumed there is no space-time coding in the NDA case. The effect of antenna correlation and space-time coding will be examined later. The effect of the number of transmit antennas N is shown in Figures 7 and 8 for the DA and NDA cases respectively, with $M = 4$. From both figures, it can be seen that different numbers of transmit antennas result in similar estimation accuracies. Therefore, the MSEs are approximately independent of N for both ML_{DA} and ML_{NDA}. Next, the effect of the number of receive antennas M is shown in Figures 9 and 10 for DA and

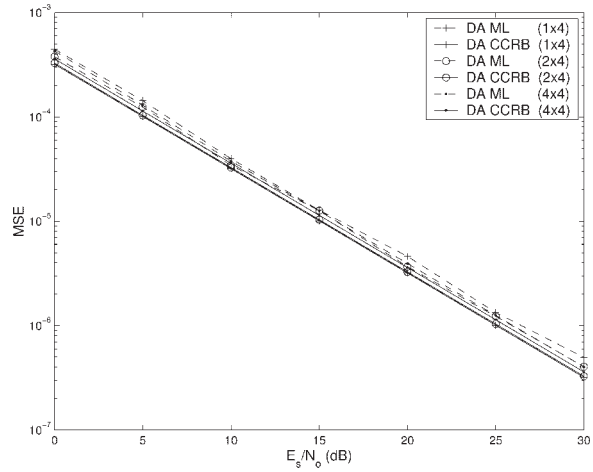

Fig. 7. MSEs of the ML_{DA} estimator and the corresponding CCRBs with different number of transmit antennas ($\mathbf{\Phi}_T = \mathbf{I}_N$, $\mathbf{\Phi}_R = \mathbf{I}_M$, $\mathbf{Z} = \mathbf{Z}_{\text{opt}}$).

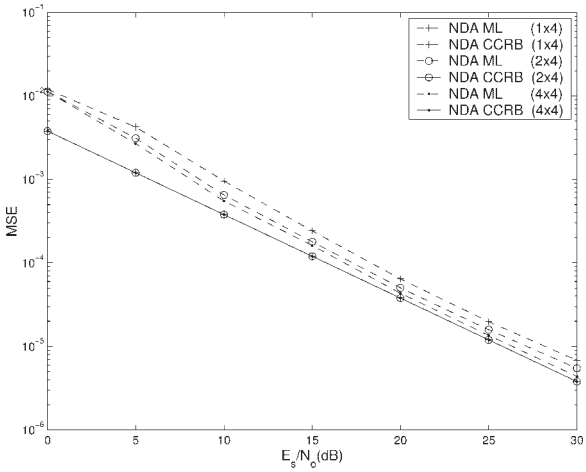

Fig. 8. MSEs of the non-data aided maximum likelihood $\mathrm{ML}_{\mathrm{NDA}}$ estimator and the corresponding CCRBs with different number of transmit antennas ($\mathbf{\Phi}_T = \mathbf{I}_N$, $\mathbf{\Phi}_R = \mathbf{I}_M$ and the data transmitted is spatially and temporally white).

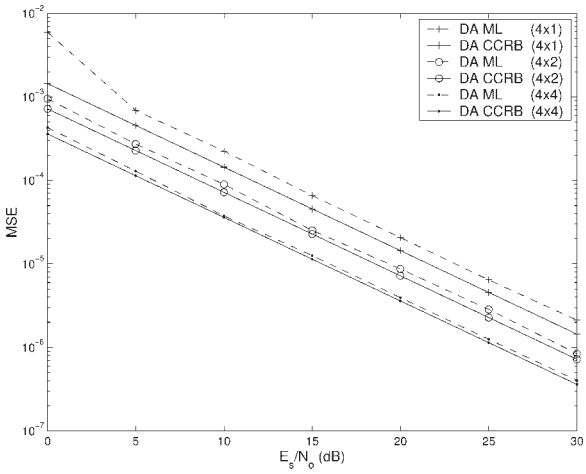

Fig. 9. MSEs of the $\mathrm{ML}_{\mathrm{DA}}$ estimator and the corresponding CCRBs with different number of receive antennas ($\mathbf{\Phi}_T = \mathbf{I}_N$, $\mathbf{\Phi}_R = \mathbf{I}_M$, $\mathbf{Z} = \mathbf{Z}_{\mathrm{opt}}$).

NDA case respectively, with $N = 4$. It is clear that increasing M leads to considerable MSE improvements. Since, from Equations (32) and (51), the $\mathrm{CCRB}_{\mathrm{DA}}$ and $\mathrm{CCRB}_{\mathrm{NDA}}$ are inversely proportional to M and from Figures 9 and 10, the performances of $\mathrm{ML}_{\mathrm{DA}}$ and $\mathrm{ML}_{\mathrm{NDA}}$ are very close to their corresponding CCRBs, it can be concluded that the MSEs of $\mathrm{ML}_{\mathrm{DA}}$ and $\mathrm{ML}_{\mathrm{NDA}}$ estimators are approximately inversely proportional to M.

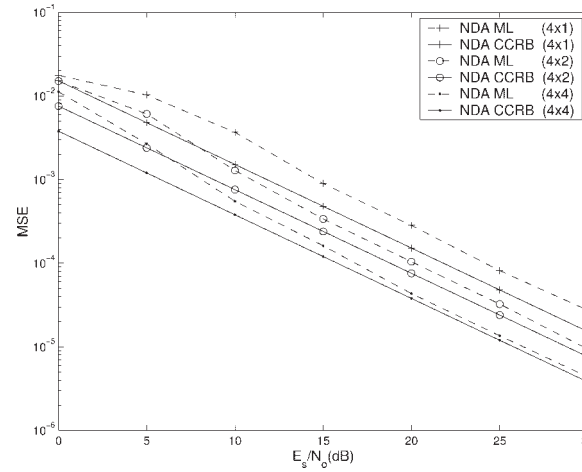

Fig. 10. MSEs of the $\mathrm{ML}_{\mathrm{NDA}}$ estimator and the corresponding CCRBs with different number of receive antennas ($\mathbf{\Phi}_T = \mathbf{I}_N$, $\mathbf{\Phi}_R = \mathbf{I}_M$ and the data transmitted is spatially and temporally white).

It is reasonable to have improved performances when the number of receive antennas increases since, more receive antennas provide diversity gain. It is tempted to argue that using more transmit antennas, should also improve the performances of symbol timing estimation since from the experience of STBC [2,5], more transmit antennas also provide diversity gain. However, notice that the diversity gain of STBC does not come automatically by just increasing the number of transmit antennas. In STBC, the observation length for demodulating a symbol has to be increased with the number of transmit antennas. For symbol timing estimation, irrespective of the number of transmit antennas, the total transmit power and the observation length are kept constant, it is not unreasonable to have MSE performances approximately independent of N. For multiple receive antennas, although the observation length (for each receive antenna) is kept constant, the observations from different receive antennas are independent (similar to the situation of maximal-ratio receive combining scheme). These independent observations increase the *effective* observation length and performance is improved due to the longer effective observation.

5.2. Effects of Correlation Among Antennas

Figures 11 and 12 show the MSE performances of $\mathrm{ML}_{\mathrm{DA}}$ and $\mathrm{ML}_{\mathrm{NDA}}$ of a 4×4 system under the effect of correlated fading among antennas. The measured

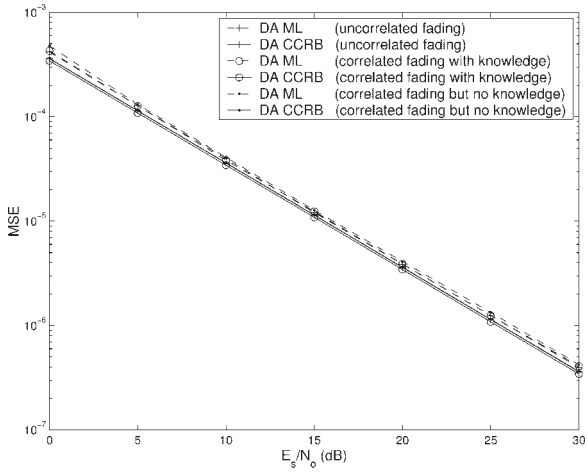

Fig. 11. MSEs of the $\mathrm{ML_{DA}}$ estimator and the corresponding CCRBs with and without fading correlation between antennas for a 4×4 system.

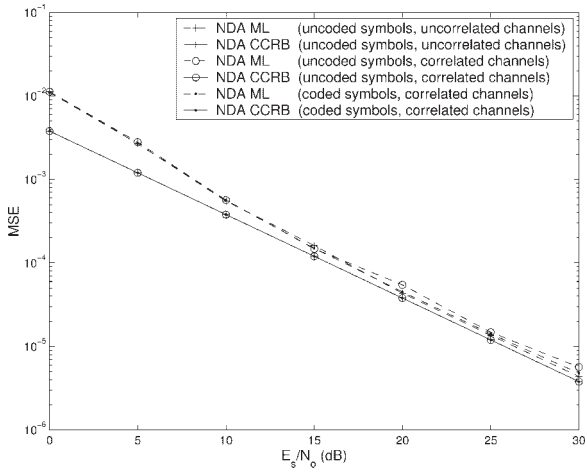

Fig. 12. MSEs of the $\mathrm{ML_{NDA}}$ estimator and the corresponding CCRBs with and without fading correlation between antennas for a 4×4 system.

correlation matrices from Nokia [10] are used in simulations

$$\mathbf{\Phi}_T = \begin{bmatrix} 1 & 0.4154 & 0.2057 & 0.1997 \\ 0.4154 & 1 & 0.3336 & 0.3453 \\ 0.2057 & 0.3336 & 1 & 0.5226 \\ 0.1997 & 0.3453 & 0.5226 & 1 \end{bmatrix},$$

$$\mathbf{\Phi}_R = \begin{bmatrix} 1 & 0.3644 & 0.0685 & 0.3566 \\ 0.3644 & 1 & 0.3245 & 0.1848 \\ 0.0685 & 0.3245 & 1 & 0.3093 \\ 0.3566 & 0.1848 & 0.3093 & 1 \end{bmatrix} \tag{60}$$

Three cases are considered in Figure 11 for the DA case. The first case assumes no correlation among antenna arrays and serves as a reference and is shown by the '+' markers. The second one, which is shown by 'o' markers, assumes that correlations exist among antennas and perfect knowledge of $\mathbf{\Phi}_T$ is available for designing optimal training sequences. The last case, denoted by the '.' markers, assumes that correlations exist among antennas but no knowledge of correlations is assumed when designing the training sequences. It can be seen that the fading correlations among antennas do not change the MSE performance of the $\mathrm{ML_{DA}}$ estimator or the $\mathrm{CCRB_{DA}}$. Furthermore, surprisingly, the knowledge of $\mathbf{\Phi}_T$ for designing optimal training sequences is not important as the results show that training sequences assuming no correlation perform equally well in the presence of correlation among antennas. For the NDA case (Figure 12), three cases are considered, too. The first one is no space-time coding and no fading correlation, which is shown using '+' markers. The second one is no space-time coding but with fading correlation, which is shown by 'o' markers. The final one is that the data is encoded with the half rate STBC (58) and with correlated fading, which is shown by '.' markers. It can be seen that the presence of correlated fading and space-time coding do not affect the MSE performances of the $\mathrm{ML_{NDA}}$ estimator.

In order to investigate the performance of $\mathrm{ML_{DA}}$ and $\mathrm{ML_{NDA}}$ estimators under different degree of fading correlation, we employ the following single parameter correlation model:

$$[\mathbf{\Phi}_T]_{ij} = [\mathbf{\Phi}_R]_{ij} = \rho^{|i-j|} \tag{61}$$

where, $\rho \in [0, 1)$ is the correlation coefficient between adjacent antennas (note that $\rho = 0$ means no correlation). Figure 13 shows the MSEs of the $\mathrm{ML_{DA}}$ estimator against ρ for $E_s/N_o = 10$, 20 and 30 dB in a 4×4 system. Two cases are considered. The first one assumes perfect knowledge of correlation for designing training sequences and the second one assumes no correlation when designing training sequences. It can be seen that for $\rho \leq 0.5$, the performance degradation due to antenna correlation is extremely small. Only when $\rho > 0.5$, the performance start to degrade, but with limited degree. Also, designing training sequences without knowledge of correlation results only in a slight degradation with respect to the case, which assumes perfect knowledge of correlation, and this only happens when $\rho > 0.5$. This property facilitates the practical implementation of

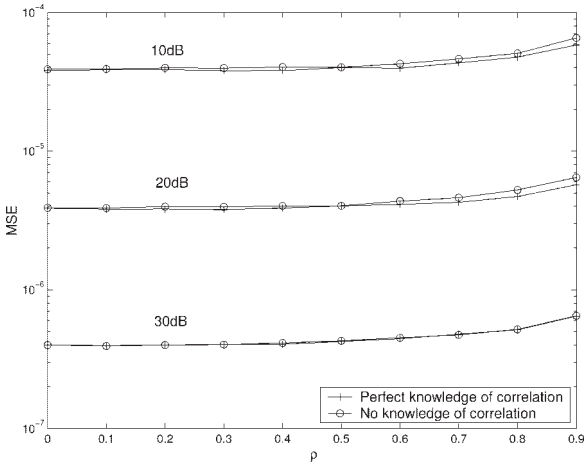

Fig. 13. MSEs of the $\mathrm{ML}_{\mathrm{DA}}$ estimator against the correlation coefficient ρ between adjacent antennas for $E_s/N_o = 10$, 20 and 30 dB in a 4×4 system.

the proposed scheme since in practice, the correlation matrix may not be perfectly known. This also explains the results in Figure 11 that the $\mathrm{ML}_{\mathrm{DA}}$ estimator does not suffer any loss of performance since the largest measured correlation coefficient between adjacent antennas in Equation (60) is about 0.5. Figure 14 shows the MSEs of the $\mathrm{ML}_{\mathrm{NDA}}$ estimator against ρ for $E_s/N_o = 10$, 20 and 30 dB in a 4×4 system. Two cases are simulated. The first case is no space-time coding, while the second case is encoded by Equation (58). It can be seen that, basically, the space-time coding considered in this example does not have any effect on the MSE performances of the $\mathrm{ML}_{\mathrm{NDA}}$ with respect to the no coding case. Furthermore, the degradation due to extreme antenna correlations is very small.

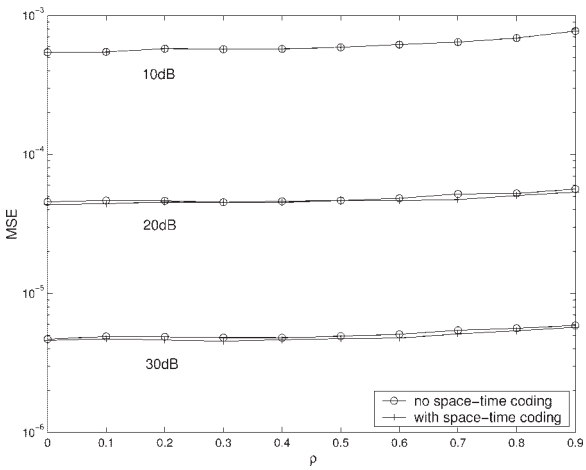

Fig. 14. MSEs of the $\mathrm{ML}_{\mathrm{NDA}}$ estimator against the correlation coefficient between adjacent antennas ρ for $E_s/N_o = 10$, 20 and 30 dB in a 4×4 system.

The small dependence of the MSEs on correlation between antennas is due to the fact that, in this study, the nuisance parameters (i.e. $\mathrm{vec}(\mathbf{H}_{\mathrm{i.i.d.}}^T)$ for DA case and $\mathrm{vec}(\mathbf{Z}\sqrt{\boldsymbol{\Phi}_T}\mathbf{H}_{\mathrm{i.i.d.}}^T)$ for NDA case) are treated as deterministic unknown and are being jointly estimated together with ε_o. The correlation between antennas, can always be lumped into the nuisance parameters. Since, this action does not change the dimension of the nuisance parameters and there is no constraint on the value of the nuisance parameters, the effect of correlation between antennas on the MSE of $\hat{\varepsilon}$ would be very small.

5.3. Comparison of DA and NDA Estimators

Here, we compare the performance of the $\mathrm{ML}_{\mathrm{DA}}$ and $\mathrm{ML}_{\mathrm{NDA}}$ estimators with their corresponding CCRBs and MCRBs for a 4×4 system. For simplicity, it is assumed that there is no correlation among antennas and there is no space-time coding for NDA case (since the effects of these are small as shown earlier). Figure 15 shows the results. Note that from Figure 15, the MSE performances of $\mathrm{ML}_{\mathrm{DA}}$ and $\mathrm{ML}_{\mathrm{NDA}}$ estimators are very close to their corresponding CCRBs. This means that $\mathrm{ML}_{\mathrm{DA}}$ and $\mathrm{ML}_{\mathrm{NDA}}$ are efficient estimators conditioned that the nuisance parameters are being jointly estimated together with the unknown timing delay. Also, note that the performance of $\mathrm{ML}_{\mathrm{DA}}$ estimator is very close to the $\mathrm{MCRB}_{\mathrm{DA}}$, which implies that $\mathrm{ML}_{\mathrm{DA}}$ is almost the best possible estimator under the problem at hand, regardless of how we deal with the nuisance parameters. For the NDA

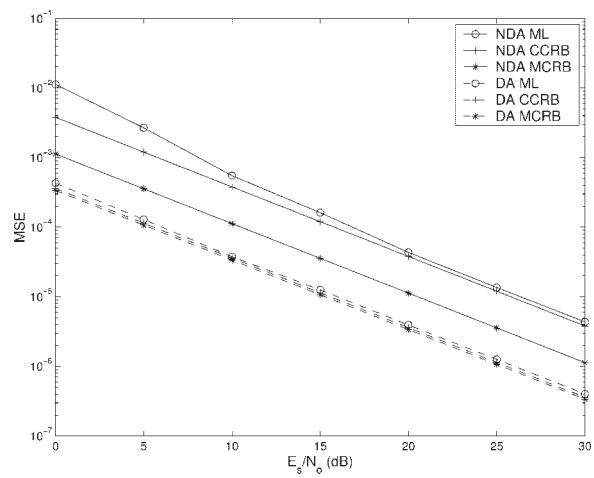

Fig. 15. Comparison of MSEs of the $\mathrm{ML}_{\mathrm{NDA}}$ and $\mathrm{ML}_{\mathrm{DA}}$ and their corresponding CCRBs and MCRBs for a 4×4 system.

case, unfortunately, although the performance of $\mathrm{ML_{NDA}}$ estimator reaches the corresponding $\mathrm{CCRB_{NDA}}$, the $\mathrm{CCRB_{NDA}}$ is quite far away from the $\mathrm{MCRB_{NDA}}$. Notice that, according to Reference [21], CCRB is a valid bound only for estimators that rely on quadratic non-linearity, there is a possibility that some other NDA estimators employing higher order (>2) non-linearities would have performances closer to the MCRB. This is subject to further investigations.

Finally, as expected, $\mathrm{ML_{DA}}$ estimator performs much better than the $\mathrm{ML_{NDA}}$ estimator. However, this comes with a price. The $\mathrm{ML_{DA}}$ estimator requires training sequences, resulting in lower transmission efficiency. Moreover, the estimation has to be performed at specific times when the training data is available, while $\mathrm{ML_{NDA}}$ can be performed at any time during transmission. This also means that, for the DA case, there is a need to synchronize the training sequences before timing estimation. This requires extra implementation complexity. In addition, degradation may occur if the positions of the training sequences are mislocated. Furthermore, the computation of the DA likelihood function (20) is more complicated than that of the NDA likelihood function (43). Therefore, $\mathrm{ML_{DA}}$ and $\mathrm{ML_{NDA}}$ provide a performance, transmission efficiency and complexity tradeoff for symbol timing estimation in MIMO channels.

6. Conclusions

The DA and NDA ML symbol timing estimators, their corresponding CCRB and MCRB for MIMO correlated flat-fading channels have been derived in this paper. For the DA case, the optimal orthogonal training sequences have also been derived. It was shown that the approximated ML algorithm in References [4,13] is just a special case of the DA ML algorithm; while the extended squaring algorithm in Reference [14] is just a special case of the NDA ML estimator. For the optimal orthogonal training sequences, it was found that they resemble Walsh sequences but with modified envelopes. Simulation results under different operating conditions (e.g. number of antennas and correlation between antennas) were given to assess the performances of the DA and NDA ML estimators and compare them with the corresponding CCRBs and MCRBs. It was found that (i) the MSE of the DA ML estimator is close to the CCRB and MCRB, meaning that the DA ML estimator is almost the best estimator (in terms of MSE performance) for the problem under consideration, (ii) the MSE of the NDA ML estimator is close to the CCRB but not MCRB, meaning that NDA ML estimator is an efficient estimator conditioned that the nuisance parameters are being jointly estimated, but there might exist other NDA estimators with better performances, (iii) the MSEs of both DA and NDA ML estimators are approximately independent of the number of transmit antennas and are inversely proportional to the number of receive antennas, (iv) correlation between antennas has little impact on the MSEs of DA and NDA ML estimators unless the correlation coefficient between adjacent antennas is larger than 0.5, in which case a small degradation occurs and (vi) DA ML performs better than NDA ML estimator at the cost of lower transmission efficiency and higher implementation complexity.

Appendix I: Proof of Equation (35)

From the expression of $\mathrm{MCRB_{DA}}$ in Equation (34), only the product inside the $\mathrm{tr}(.)$ operator depends on $\mathbf{Z}$, therefore the problem of finding optimal training sequence is equivalent to maximizing $\mathrm{tr}(\tilde{\mathbf{Z}}^H\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}\tilde{\mathbf{Z}}\mathbf{\Phi}_T)$ with respect to $\mathbf{Z}$ with the constraints that (i) the columns of $\mathbf{Z}$ have to be independent of each other and (ii) $[\mathbf{Z}^H\mathbf{Z}]_{ii} = L_o + 2L_g$ for $i = 1, \ldots, N$. The first constraint is for the $\mathrm{ML_{DA}}$ to hold and has been mentioned before. The second constraint is the power constraint, and we assume that the training sequence has average unit energy on each data bit. Now, consider the eigenvector decomposition $\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o} = \mathbf{U}_D\mathbf{\Sigma}_D\mathbf{U}_D^H$, where, $\mathbf{\Sigma}_D$ is a diagonal matrix with the eigenvalues of $\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}$ located on the diagonal and $\mathbf{U}_D$ is the unitary matrix containing all the corresponding eigenvectors as columns. Similarly, express $\mathbf{\Phi}_T = \mathbf{U}_T\mathbf{\Sigma}_T\mathbf{U}_T^H$. Then

$$\mathrm{tr}\left(\tilde{\mathbf{Z}}^H\tilde{\mathbf{D}}_{\varepsilon_o}^H\tilde{\mathbf{D}}_{\varepsilon_o}\tilde{\mathbf{Z}}\mathbf{\Phi}_T\right) = \mathrm{tr}\left(\tilde{\mathbf{Z}}^H\mathbf{U}_D\mathbf{\Sigma}_D\mathbf{U}_D^H\tilde{\mathbf{Z}}\mathbf{U}_T\mathbf{\Sigma}_T\mathbf{U}_T^H\right) \tag{62}$$

$$= \mathrm{tr}\left(\sqrt{\mathbf{\Sigma}_T}^H\mathbf{U}_T^H\tilde{\mathbf{Z}}^H\mathbf{U}_D\mathbf{\Sigma}_D\mathbf{U}_D^H\tilde{\mathbf{Z}}\mathbf{U}_T\sqrt{\mathbf{\Sigma}_T}\right) \tag{63}$$

$$= \mathrm{tr}\left(\Xi^H\mathbf{\Sigma}_D\Xi\right) \tag{64}$$

$$= \sum_{i=1}^{N}[\Xi_{:,i}]^H\mathbf{\Sigma}_D\Xi_{:,i} \tag{65}$$

where, $\Xi \triangleq \mathbf{U}_D^H\tilde{\mathbf{Z}}\mathbf{U}_T\sqrt{\mathbf{\Sigma}_T}$. Note that, if we set $\mathbf{Z}^H\mathbf{Z} = (L_o + 2L_g)\mathbf{I}_N$ (this is a sufficient condition that makes the two constraints mentioned earlier

satisfied), then the columns of Ξ are orthogonal to each other (since $\Xi^H\Xi = \sqrt{\boldsymbol{\Sigma}_T}^H \mathbf{U}_T^H \tilde{\mathbf{Z}}^H \mathbf{U}_D \ \mathbf{U}_D^H \tilde{\mathbf{Z}} \mathbf{U}_T \sqrt{\boldsymbol{\Sigma}_T} = (L_o + 2L_g)\boldsymbol{\Sigma}_T/N$). Therefore, by confining the training sequences to be orthogonal, the problem then becomes to maximize $[\Xi_{:,i}]^H \boldsymbol{\Sigma}_D \Xi_{:,i}$ with respect to $\Xi_{:,i}$ for each i with the constraints that $[\Xi_{:,i}]^H \Xi_{:,i} = (L_o + 2L_g)[\boldsymbol{\Sigma}_T]_{ii}/N$ and $[\Xi_{:,i}]^H \Xi_{:,j} = 0$ for $j = 1, \ldots, i-1$.

It is well-known that for a Hermitian matrix $\mathbf{R}$, the vector $\mathbf{u}$ that maximizes $\mathbf{u}^H\mathbf{R}\mathbf{u}$ subject to the constraints that $\|\mathbf{u}\| = 1$ and $\mathbf{u}^H\mathbf{u}_i = 0$, for $i = 1, 2, \ldots, k-1$, where $\mathbf{u}_i$ is the eigenvector corresponding to the ith largest eigenvalue of $\mathbf{R}$, is $\mathbf{u}_k$ [16]. Setting $\mathbf{R} = \boldsymbol{\Sigma}_D$ and with the proper power constraints, it is not difficult to see that $\Xi_{:,i}$ is the eigenvector corresponding to the ith largest eigenvalue of $\boldsymbol{\Sigma}_D$ scaled by the energy factor $\sqrt{(L_o + 2L_g)[\boldsymbol{\Sigma}_T]_{ii}/N}$. Since $\boldsymbol{\Sigma}_D$ is a diagonal matrix, the ith eigenvector is a vector of length $L_o + 2L_g$ with one at the ith position and zero at other positions. Therefore,

$$\Xi = \sqrt{\frac{(L_o + 2L_g)}{N}} \begin{bmatrix} \sqrt{\boldsymbol{\Sigma}_T} \\ \mathbf{0}_{(L_o+2L_g-N)\times N} \end{bmatrix} \tag{66}$$

where, $\mathbf{0}_{(L_o+2L_g-N)\times N}$ is an all zero matrix with dimensions $(L_o + 2L_g - N) \times N$, with $\Xi = \mathbf{U}_D^H \tilde{\mathbf{Z}} \mathbf{U}_T \sqrt{\boldsymbol{\Sigma}_T}$, we have

$$\begin{aligned} \mathbf{Z} &= \sqrt{(L_o + 2L_g)}\mathbf{U}_D \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{(L_o+2L_g-N)\times N} \end{bmatrix} \\ \mathbf{U}_T^H &= \sqrt{(L_o + 2L_g)}\tilde{\mathbf{U}}(\varepsilon_o)\mathbf{U}_T^H \end{aligned} \tag{67}$$

where, $\tilde{\mathbf{U}}(\varepsilon_o)$ is the matrix containing the N eigenvectors corresponding to the N largest eigenvalues of $\tilde{\mathbf{D}}_{\varepsilon_o}^H \tilde{\mathbf{D}}_{\varepsilon_o}$ as columns.

Appendix II: Proof of Equation (49)

First note that $\boldsymbol{\Gamma}_\mathbf{x}$ can be rewritten in the following form:

$$\begin{aligned} \boldsymbol{\Gamma}_\mathbf{x} &\triangleq \mathbb{E}\left[\mathrm{vec}\left(\mathbf{Z}\sqrt{\boldsymbol{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T\right)\mathrm{vec}\left(\mathbf{Z}\sqrt{\boldsymbol{\Phi}_T}\mathbf{H}_{\text{i.i.d.}}^T\right)^H\right] \\ &= \mathbb{E}\left[(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})\mathrm{vec}\left(\sqrt{\boldsymbol{\Phi}_T}\right)\mathrm{vec}\left(\sqrt{\boldsymbol{\Phi}_T}\right)^H(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})^H\right] \\ &= \mathbb{E}\left[(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)\right] \end{aligned} \tag{68}$$

where, $\Upsilon \triangleq \mathrm{vec}(\sqrt{\boldsymbol{\Phi}_T})\mathrm{vec}(\sqrt{\boldsymbol{\Phi}_T})^H$. The (i,j)th element of $\boldsymbol{\Gamma}_\mathbf{x}$ is given by

$$\begin{aligned} [\boldsymbol{\Gamma}_\mathbf{x}]_{ij} &= \mathbb{E}\left[(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:}\Upsilon(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}\right] \\ &= \mathbb{E}\left[\mathrm{tr}\left(\Upsilon(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:}\right)\right] \\ &= \mathrm{tr}\left(\Upsilon\mathbb{E}[(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:}]\right) \end{aligned} \tag{69}$$

with $i, j = 0, 1, \ldots, M(L_o + 2L_g)$.

Let $i = i_q(L_o + 2L_g) + i_r$ and $j = j_q(L_o + 2L_g) + j_r$ such that $i_q, j_q \in \{0, 1, \ldots, M-1\}$ and $i_r, j_r \in \{1, \ldots, L_o + 2L_g\}$ are the quotients and remainders of divisions $i/(L_o + 2L_g)$ and $j/(L_o + 2L_g)$ respectively. Also

$$\begin{aligned} &\mathbb{E}[(\mathbf{H}_{\text{i.i.d.}}^H \otimes \mathbf{Z}^H)_{:,j}(\mathbf{H}_{\text{i.i.d.}} \otimes \mathbf{Z})_{i,:}] \\ &\quad = \mathbb{E}\left[\left((\mathbf{H}_{\text{i.i.d.}}^H)_{:,j_q} \otimes (\mathbf{Z}^H)_{:,j_r}\right)\left((\mathbf{H}_{\text{i.i.d.}})_{i_q,:} \otimes (\mathbf{Z})_{i_r,:}\right)\right] \\ &\quad = \mathbb{E}\left[(\mathbf{H}_{\text{i.i.d.}}^*)_{j_q,:}^T(\mathbf{H}_{\text{i.i.d.}})_{i_q,:}] \otimes \mathbb{E}[(\mathbf{Z}^*)_{j_r,:}^T(\mathbf{Z})_{i_r,:}\right] \\ &\quad = \mathbf{I}_N\delta_{i_qj_q} \otimes \boldsymbol{\Gamma}_z(j_r - i_r) \end{aligned} \tag{70}$$

where, $\boldsymbol{\Gamma}_z(j_r - i_r) \triangleq \mathbb{E}[(\mathbf{Z}^*)_{j_r,:}^T(\mathbf{Z})_{i_r,:}]$ is the average cross-correlation matrix of the symbols transmitted with the time index difference $j_r - i_r$. Note that $\mathbb{E}[(\mathbf{Z}^*)_{j_r,:}^T(\mathbf{Z})_{i_r,:}]$ depends only on the time index difference but not on the absolute time index since, in the NDA case we never know where the observation begins, the average cross-correlation between time indices 1 and 3 would be the same as that for time indices 5 and 7. Putting Equation (70) into Equation (69), we obtain

$$[\boldsymbol{\Gamma}_\mathbf{x}]_{ij} = \delta_{i_qj_q}\mathrm{tr}(\Upsilon(\mathbf{I}_N \otimes \boldsymbol{\Gamma}_z(j_r - i_r))) \tag{71}$$

implying that

$$\boldsymbol{\Gamma}_\mathbf{x} = \mathbf{I}_M \otimes \boldsymbol{\Psi} \tag{72}$$

where, $\boldsymbol{\Psi}$ is a Hermitian, Toeplitz matrix with $[\boldsymbol{\Psi}]_{ij} = \mathrm{tr}(\Upsilon(\mathbf{I}_N \otimes \boldsymbol{\Gamma}_z(j - i)))$. Note that $[\boldsymbol{\Psi}]_{ij}$ can be simplified as

$$\begin{aligned} [\boldsymbol{\Psi}]_{ij} &= \mathrm{tr}\left(\mathrm{vec}\left(\sqrt{\boldsymbol{\Phi}_T}\right)\mathrm{vec}\left(\sqrt{\boldsymbol{\Phi}_T}\right)^H(\mathbf{I}_N \otimes \boldsymbol{\Gamma}_z(j - i))\right) \\ &= \mathrm{tr}\left(\mathrm{vec}\left(\sqrt{\boldsymbol{\Phi}_T}\right)^H(\mathbf{I}_N \otimes \boldsymbol{\Gamma}_z(j - i))\mathrm{vec}\left(\sqrt{\boldsymbol{\Phi}_T}\right)\right) \\ &= \mathrm{tr}\left(\left(\sqrt{\boldsymbol{\Phi}_T}\right)^H\boldsymbol{\Gamma}_z(j - i)\left(\sqrt{\boldsymbol{\Phi}_T}\right)\right) \\ &= \mathrm{tr}(\boldsymbol{\Gamma}_z(j - i)\boldsymbol{\Phi}_T) \end{aligned} \tag{73}$$

Appendix III: Proof of Equation (57)

First note that the observation interval usually involves more than one independent space-time encoded block, each given by the form (56), therefore $\mathbf{\Gamma}_z(j-i) = \mathbf{0}_N$ for $|j-i| \geq s$. Furthermore, since $\mathbf{\Gamma}_z(j-i) = \mathbf{\Gamma}_z{}^*(i-j)$, it is sufficient to concentrate on $\mathbf{\Gamma}_z(j-i)$, for $j-i=\ell$ with $\ell = 0, 1, \ldots, s-1$

$$\mathbf{\Gamma}_z(\ell) = \frac{1}{s}\sum_{n=1}^{s-\ell} \mathbb{E}[(\mathcal{G}_{n+\ell,:})^H(\mathcal{G}_{n,:})] \tag{74}$$

where, the factor $1/s$ exists because in NDA estimation, the probability that the observation start at a particular row of the matrix $\mathcal{G}$ is $1/s$. Putting Equation (56) into Equation (74), we obtain

$$\begin{aligned}\mathbf{\Gamma}_z(\ell) = \frac{1}{s}\sum_{n=1}^{s-\ell} \mathbb{E}\Bigg[&\left(\sum_{k=1}^{rs} \Re(b_k)\mathbf{X}_k + j\sum_{k=1}^{rs} \Im(b_k)\mathbf{Y}_k\right)^H_{n+\ell,:} \\ &\left(\sum_{k'=1}^{rs} \Re(b_{k'})\mathbf{X}_{k'} + j\sum_{k'=1}^{rs} \Im(b_{k'})\mathbf{Y}_{k'}\right)_{n,:}\Bigg] \\ = \frac{1}{s}\sum_{n=1}^{s-\ell} \Bigg(&\sum_{k=1}^{rs} \mathbb{E}[\Re(b_k)\Re(b_k)][\mathbf{X}_k]^T_{n+\ell,:}[\mathbf{X}_k]_{n,:} \\ &+ \sum_{k=1}^{rs} \mathbb{E}[\Im(b_k)\Im(b_k)][\mathbf{Y}_k]^T_{n+\ell,:}[\mathbf{Y}_k]_{n,:}\Bigg)\end{aligned} \tag{75}$$

where, we have used the i.i.d. property of b_k, $\mathbb{E}[\Re(b_k)\Re(b_{k'})] = 0$, $\mathbb{E}[\Im(b_k)\Im(b_{k'})] = 0$ for $k \neq k'$ and $\mathbb{E}[\mathcal{R}(b_k)\Im(b_{k'})] = 0$ for all combination of k and k'. Further note that, $\mathbb{E}[\Re(b_k)\Re(b_k)] = \mathbb{E}[\Im(b_k)\Im(b_k)] = 1/2$, then we have for $j-i=\ell$ with $\ell = 0, 1, \ldots, s-1$

$$\begin{aligned}\mathbf{\Gamma}_z(j-i) = \frac{1}{2s}\sum_{n=1}^{s-\ell} \Bigg(&\sum_{k=1}^{rs} [\mathbf{X}_k]^T_{n+\ell,:}[\mathbf{X}_k]_{n,:} \\ &+ \sum_{k=1}^{rs} [\mathbf{Y}_k]^T_{n+\ell,:}[\mathbf{Y}_k]_{n,:}\Bigg)\end{aligned} \tag{76}$$

Finally, note that since $\mathbf{X}_k$ and $\mathbf{Y}_k$ are real-valued, $\mathbf{\Gamma}_z(j-i)$ would also be real-valued and $\mathbf{\Gamma}_z(j-i) = \mathbf{\Gamma}_z(i-j)$. Therefore, it can be concluded that Equation (76) is true for $|j-i| = \ell$ ($\ell = 0, 1, \ldots, s-1$).

References

1. Foschini GJ, Gans MJ. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications* 1998; **6**: 311–335.
2. Alamouti SM. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications* 1998; **16**: 1451–1458.
3. Naguib AF, Seshadri N, Calderbank AR. Increasing data rate over wireless channels. *IEEE Signal Processing Magazine* 2000; **17**: 76–92.
4. Naguib AF, Tarokh V, Seshadri N, Calderbank AR. A space-time coding modem for high-data-rate wireless communications. *IEEE Journal on Selected Areas in Communications* 1998; **16**: 1459–1478.
5. Tarokh V, Jafarkhani H, Calderbank AR. Space-time block coding for wireless communications: performance results. *IEEE Journal on Selected Areas in Communications* 1999; **17**: 451–460.
6. Larsson EG, Stoica P, Li J. Orthogonal space-time block codes: maximum likelihood detection for unknown channels and unstructured interferences. *IEEE Transactions on Signal Processing* 2003; **51**: 362–372.
7. Yu K, Bengtsson M, Ottersten B, McNamara D, Karlsson P, Beach M. Modeling of wide-band MIMO radio channels based on NLoS indoor measurements. *IEEE Transactions on Vehicular Technology* 2004; **53**: 655–665.
8. Kermoal JP, Schumacher L, Pedersen KI, Mogensen PE, Frederisken F. A stochastic MIMO radio channel model with experimental validation. *IEEE Journal on Selected Areas in Communications* 2002; **20**: 1211–1226.
9. Chizhik D, Ling J, Wolniansky PW, Valenzuela RA, Costa N, Huber K. Multiple-input-multiple-output measurements and modeling in Manhattan. *IEEE Journal on Selected Areas in Communications* 2003; **21**: 321–331.
10. Schumacher L, Kermoal JP, Frederiksen F, Pedersen KI, Algrans A, Mogensen PE. MIMO channel characterisation. *Deliverable D2 V1.1 of IST-1999-11729 METRA Project*, pp. 1–57, February 2001. Available online: http://www.ist-metra.org
11. Gesbert D, Sha M, Shiu DS, Smith PJ, Naguib A. From theory to practice: an overview of MIMO space-time coded wireless systems. *IEEE Journal on Selected Areas in Communications* 2003; **21**: 281–301.
12. Foschini GJ, Golden GD, Valenzuela RA, Wolniansky PW. Simplified processing for high spectral efficiency wireless communication employing multi-element arrays. *IEEE Journal on Selected Areas Communications* 1999; **17**: 1841–1852.
13. Wu YC, Chan SC, Serpedin E. Symbol-timing estimation in space-time coding systems based on orthogonal training sequences, accepted for publication in *IEEE Transactions on Wireless Communications* 2004. Available online: http://ee.tamu.edu/∼serpedin
14. Wu YC, Chan SC. On the symbol timing recovery in space-time coding systems. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)* March 2003; pp. 420–424.
15. Kay SM. *Fundamentals of Statistical Signal Processing—Estimation Theory*. Prentice Hall: New Jersey, 1993.
16. Horn RA, Johnson CR. *Matrix analysis*. Cambridge University Press: New York, 1990.
17. Zakharov YV, Baronkin VM, Tozer TC. DFT-based frequency estimators with narrow acquisition range. *IEE Proceedings-Communications* 2001; **148**(1): 1–7.
18. Stoicia P, Nehorai A. MUSIC, maximum likelihood and Cramer–Rao bound. *IEEE Transactions Acoustics speech, signal processing* 1989; **37**: 720–741.

19. Ottersten B, Viberg M, Stoica P, Nehorai A. Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing. In *Radar Array Processing*. Springer-Verlag: New York, 1993.
20. D'Andrea AN, Mengali U, Reggiannini R. The modified Cramer–Rao bound and its application to synchronization problem. *IEEE Transactions Communications* 1994; **42**: 1391–1399.
21. Vazquez G, Riba J. Non-data-aided digital synchronization. In *Signal Processing Advanced in Wireless and Mobile Communications*, Vol. 2, Giannakis GB, Hua Y, Stoica P, Tong L (eds). Prentice Hall: New Jersey, 2001.
22. Riba J, Sala J, Vazquez G. Conditional maximum likelihood timing recovery: estimators and bounds. *IEEE Transactions on Signal Processing* 2001; **49**: 835–850.
23. Wu YC, Serpedin E. Low-complexity feedforward symbol timing estimator using conditional maximum likelihood principle. *IEEE Communications Letters* 2004; **8**: 168–170.
24. Oerder M, Meyr H. Digital filter and squaring timing recovery. *IEEE Transactions Communications* 1988; **36**: 605–612.

Authors' Biographies

Yik-Chung Wu received his B.Eng. (honors) and M.Phil. in electronic engineering from the University of Hong Kong in 1998 and 2001 respectively. He was then a research assistant in the same university from 2001 to 2002. He obtained the Croucher Foundation scholarship in 2002 and currently, he is pursing his Ph.D. at the Texas A&M University, College Station, TX, U.S.A. His research interests include digital signal processing with applications to communication systems, software radio and space-time processing.

Erchin Serpedin received (with highest distinction) the diploma of electrical engineering from the Polytechnic Institute of Bucharest, Bucharest, Romania, in 1991. He received the specialization degree in signal processing and transmission of information from Ecole Superiéure D'Electricité, Paris, France, in 1992, the M.Sc. degree from Georgia Institute of Technology, Atlanta, GA, in 1992, and the Ph.D. in electrical engineering from the University of Virginia, Charlottesville, VA, in January 1999. From 1993 to 1995, he was an instructor in the Polytechnic Institute of Bucharest, and between January and June 1999, he was a lecturer at the University of Virginia. In July 1999, he joined Texas A&M University, Wireless Communications Laboratory, as an assistant professor. His research interests lie in the areas of statistical signal processing and wireless communications. He has received the NSF Career Award in 2001, and is currently an associate editor for the *IEEE Communications Letters* and the *IEEE Signal Processing Letters*.