# Near Optimal Training Sequences for Low Complexity Symbol Timing Estimation in MIMO Systems

Ketan Rajawat, *Student Member, IEEE*, and A. K. Chaturvedi, *Senior Member, IEEE*

***Abstract*—Training sequences for data-aided timing estimation in multi-input multi-output systems are designed. It is observed that for low complexity implementation, the sequences must necessarily satisfy the zero cross-correlation zone property. By restricting our search to a more tractable subset of this class of sequences, we are able to minimize the modified Cramer-Rao bound in closed form and obtain sequences whose mean square error performance is close to that of the optimal orthogonal sequences. Two constant modulus sequences with even lower implementation complexity are also proposed.**

***Index Terms*—MIMO, symbol timing, orthogonal sequences.**

## I. INTRODUCTION

SYMBOL timing estimation plays a critical role in realizing the potential of multi-antenna systems. We consider the problem of data-aided feedforward symbol timing estimation in multi-antenna systems. Feedforward estimation involves the estimation of the optimal timing instants using the oversampled received signal through maximum likelihood (ML)-oriented arguments. The optimum sample selection algorithm was the first work to address this problem [1]. However, in order to obtain a reasonable performance, the algorithm required a large oversampling factor. Recently a Discrete Fourier Transform (DFT) based interpolation method [2] and a pulse shape based method [3] have been proposed which work well even with oversampling factors of four and two respectively.

The above methods use perfect training sequences with cyclic prefixes and suffixes as proposed in [2]. However as has been pointed out in [4] these perfect sequences are far from optimal for symbol timing estimation. In this paper we propose a new class of generalized orthogonal (GO) training sequences that achieve mean square error (MSE) performance close to that obtained by optimal orthogonal sequences, at much less implementation complexity. Two constant modulus sequences that have even lower complexity (but slightly worse performance) are also analyzed.

Paper approved by E. Serpedin, the Editor for Synchronization and Sensor Networks of the IEEE Communications Society. Manuscript received October 15, 2008; revised March 31, 2009.

K. Rajawat is with the Department of Electrical Engineering, University of Minnesota, MN (e-mail: ketan@umn.edu). He was earlier at the Department of Electrical Engineering, Indian Institute of Technology Kanpur, India.

A. K. Chaturvedi is with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, India (e-mail: akc@iitk.ac.in).

Digital Object Identifier 10.1109/TCOMM.2010.01.080537

The organization of the paper is as follows: Section II presents the system model used. Section III describes the generalized orthogonality criterion which the proposed sequences must satisfy. The training sequences are designed in Section IV. Simulation results and discussions are presented in Section V and finally Section VI concludes the paper.

## II. SYSTEM MODEL

We consider the space time code based modem of [1] with $N$ transmit and $M$ receive antennas. The baseband equivalent model and the notations are the same as in [3], [2]. Thus we assume the training sequence consists of $L_t + 2L_g$ symbols of which the first and the last $L_g$ symbols constitute the guard band and the middle $L_t$ symbols are the training symbols. The matched filtered signal at $j$th receive antenna is given by,

$$r_{oj}(t) = \sqrt{\frac{E_s}{N}} \sum_{k=1}^{N} h_{kj} \sum_{n=-L_g}^{L_t+L_g-1} c_k(n) p(t - nT - \epsilon_0 T) + \eta_j(t) \tag{1}$$

for $1 \leq j \leq M$, where $E_s/N$ is the symbol energy; $T$ is the symbol period; $h_{kj}$s are independent complex channel gains corresponding to the channel between $k$th transmit and $j$th receive antenna; $Q$ is the oversampling factor; $\epsilon_0$ is the timing error and $p(t) = g(t) \otimes g(t)$ is a Nyquist pulse with bandwidth $(1+\alpha)/2T$, $\alpha$ being the excess bandwidth factor and $g(t)$ the transmit/receive filter; the noise term $\eta_j(t) = n_j(t) \otimes g(t)$ where $n_j(t)$ is the circularly-symmetric, complex-valued and white Gaussian random process at the $j$th receive antenna, with power density $N_0$ ($\otimes$ denotes convolution); $\mathbf{s}_k = [c_k(-L_g)\ c_k(-L_g+1)\ \ldots\ c_k(L_t+L_g-1)]$ is the training sequence transmitted from the $k$th transmit antenna. Note that unlike [2], we do not restrict ourselves only to sequences with cyclic suffixes and prefixes. Substituting $t = (l+\epsilon)T$ and $\epsilon_0' = \epsilon_0 - k_0/Q$ (where $l$ is an integer and $k_0 = -\lfloor (1/2 - \epsilon_0)Q \rfloor$) in (1) we get the expression shown in (2).

Defining,

$$\mathbf{r}_{oj}(\epsilon) = \left[ r_{oj}(\epsilon T)\ r_{oj}((1+\epsilon)T)\ \ldots\ r_{oj}((L_t - 1 + \epsilon)T) \right]^T$$
$$\eta_j(\epsilon) = \left[ \eta_j(\epsilon T)\ \eta_j((1+\epsilon)T)\ \ldots\ \eta_j((L_t - 1 + \epsilon)T) \right]^T$$
$$\text{and } \mathbf{c}_i = \left[ c_i(0) \quad c_i(1)\ \ldots\ c_i(L_t - 1) \right]^T$$

for all $1 \leq i \leq N$, where the superscript $T$ denotes transpose. The approximated log-likelihood function for timing estima-

$$r_{oj}((l+\epsilon)T) = \sqrt{\frac{E_s}{N}} \sum_{k=1}^{N} h_{kj} \sum_{n=-L_g}^{L_t+L_g-1} c_k(n) p(\epsilon T + (l-n)T - \epsilon_0' T) + \eta_j(t) \quad 0 \le \epsilon \le 1 \tag{2}$$

tion is thus given by (similar to [2], [1])

$$\Lambda(\epsilon) = \sum_{i=1}^{N} \sum_{j=1}^{M} \Lambda_{ij}(\epsilon) \tag{3}$$

where $\Lambda_{ij}(\epsilon) = \left|\mathbf{c}_i^H \mathbf{r}_{oj}(\epsilon)\right|^2$ and superscript $H$ denotes conjugate transpose. Note that we have dropped all multiplicative constants from the expression of log likelihood function since they do not effect the maximization.

## III. The Generalized Orthogonality Criterion

As pointed out in [2], the aim of a symbol timing estimation block is to estimate $\epsilon_0'$ from the log likelihood function $\Lambda(\epsilon)$ by maximizing it over all possible $\epsilon \in [0,1)$. In digital receivers however, only $Q$ samples of $\Lambda(\epsilon)$ are available for estimation. We can simply choose the maximum of these samples which was the approach suggested in [1]. Some of the slightly more complex approaches are the interpolation based approach of [2] and the pulse shape based approach of [3]. These estimators have lower implementation complexity compared to maximum likelihood (ML) methods and achieve better MSE performance and lower sampling rates than [1].

The Cramer Rao Bound (CRB) which is a lower bound on the MSE of an estimator serves as an indicator of how good an estimator or a sequence is. As in this case, the calculation of CRB is usually intractable and modified CRB (MCRB) or Conditional CRB (CCRB) are generally used for evaluating the estimators. However as evident from [3, Fig. 1], the MSE performance of these estimators is still far from the MCRB. While perfect sequences (defined in [2]), which cancel out the Inter Symbol Interference terms and enable us to use a low complexity estimator are sub-optimal, the "optimal orthogonal sequences" (i.e. sequences designed in [4] which minimize the CCRB) require the use of ML estimator, which has very high complexity. We must therefore redesign the sequences such that the mean square error of the estimator decreases without significant increase in estimator complexity.

In this section we propose a sufficient condition that ensures a low complexity implementation of a timing estimator. Note however this condition does not necessarily result in the class of best possible sequences. The expression for $\Lambda_{ij}(\epsilon)$ can be expanded as shown in (4) where the aperiodic correlation function $R_{ik}(\tau) = \sum_{n=0}^{L_t-1} c_i^*(n) c_k(n+\tau)$.

Recall that for perfect sequences, $R_{ik}(\tau) = L_t \delta(i-k)\delta(\tau)$ where $\delta(.)$ stands for the Kronecker delta function and $\|\mathbf{s}_i\|^2 = L_t + 2L_g$. We call this the normalized energy constraint since it constrains the average energy per training symbol to be unity. The log likelihood function for perfect sequences thus becomes

$$\Lambda_{ij}(\epsilon) = \frac{1}{N_0} \frac{E_s}{N} |h_{ij}|^2 p^2(\epsilon T - \epsilon_0' T) + v_{ij}(\epsilon) \tag{5}$$

where

$$v_{ij}(\epsilon) = \frac{1}{N_0} \left|\mathbf{c}_i^H \eta_j\right|^2 + \frac{2}{N_0} \sqrt{\frac{E_s}{N}} p(\epsilon T - \epsilon_0' T) \Re\left[h_{ij}^* \mathbf{c}_i^H \eta_j\right]) \tag{6}$$

Thus at high SNRs, when the second term $v_{ij}(\epsilon)$ becomes negligible, the log likelihood function is of a known shape and can be maximized by oversampling and subsequent application of the DFT method outlined in [2].

We now want sequences that on one hand are not as restricting as perfect sequences (i.e. zero cross-correlation and delta function shaped auto-correlation) but on the other hand, allow us to use a low complexity estimator. One way of doing this is to use an estimator similar to [3] or [2] which requires the knowledge of the shape of the log-likelihood function at high SNRs. A simple condition which guarantees this is,

$$R_{ik}(\tau) = 0 \qquad \forall\ i \neq k \text{ and } |\tau| \le L_g \tag{7}$$

This requirement of a "zero cross-correlation zone" cancels out all the cross-channel gain terms except $h_{ij}$ in $\Lambda_{ij}(\epsilon)$ and also makes it independent of the phase of $h_{ij}$. We call this the generalized orthogonality criterion (GOC) and the sequences which satisfy it as GO (generalized orthogonal) sequences. The expression for $\Lambda_{ij}(\epsilon)$ at high SNR then becomes,

$$\Lambda_{ij}(\epsilon) = \frac{1}{N_0} \frac{E_s}{N} |h_{ij}|^2 \left| \sum_{\tau=-Lg}^{Lg} R_{ii}(\tau) p((\tau + \epsilon - \epsilon_0')T) \right|^2 \tag{8}$$

Thus, similar to the perfect sequences, the log likelihood function is known to us but for a scaling factor. We can now maximize this function over $\epsilon$ using the DFT method outlined in [2]. Compared to the case when we use optimal orthogonal sequences and an ML estimator (which requires real time matrix multiplications and maximization), the above sequences used with DFT method require only vector multiplications thus working at 10 to 100 times less complexity. For example, ignoring the pre-calculation and storage requirements of the ML method, it requires about $KMNL_t^2$ multiplications as opposed to $MNQL_t$ multiplications required by DFT method (here $K$ is the number of grid points required for the maximization of the objective function [4] and is typically 16 or 32).

$$\Lambda_{ij}(\epsilon) = \frac{1}{N_0} \left| \frac{E_s}{N} \sum_{k=1}^{N} h_{kj} \sum_{\tau=-Lg}^{Lg} R_{ik}(\tau) p((\tau + \epsilon - \epsilon_0')T) + \mathbf{c}_i^H \eta_j \right|^2 \tag{4}$$

## IV. Training Sequence Designs

The problem of sequence design can now be formulated as the following optimization problem,

$$\min_{\mathbf{c}\in\mathbf{G}} \mathbb{E}\left[\epsilon - \arg\min_{\epsilon'} \Lambda(\epsilon')\right]^2 \quad (9)$$

where $\mathbf{G}$ is the set of sequences which satisfy the GOC. At this stage, we make a few simplifying assumptions,

1) Note from [2, Eq. 24] that the expression for the MSE for this scheme is expected to be quite complicated and difficult to minimize over a general class of sequences. We therefore resort to minimization of the CRB, noting that this minimization might not necessarily result in the best possible sequences. Furthermore, as is usually the case with parameter estimators, in this case also, CRB is difficult to evaluate since it involves taking expectation of an exponential function. Instead we use the MCRB expression given by [9],

$$MCRB(\epsilon) = \frac{1}{-\mathbb{E}\left[\frac{\partial^2}{\partial\epsilon^2}\Lambda(\epsilon)\right]} \quad (10)$$

2) The sequences in $\mathbf{G}$ are specified by the quadratic equality constraints of the GOC, which makes finding a general expression for $\mathbf{x}$ difficult. We therefore restrict ourselves to subsets of $\mathbf{G}$ which are easy to handle. For example, in order to get $R_{ik}(\tau) = 0$ for all possible $i \neq k$, we choose a set of sequences for which $R_{ik}(\tau)$ can be grouped up into $N$ terms (corresponding to the $N$ antennas) that cancel on adding. One method is to associate each of these terms with powers of the $N$-th root of unity, $\omega = \sqrt[N]{1}$ since they too, cancel on adding. The next two subsections describe the construction of two such classes of sequences.

As shown later, these assumptions simplify the problem without significant loss in performance. The following subsections describe three types of sequences. While performance is the sole objective in the construction of Type I sequences, Type II sequences are constructed while imposing some practical constraints like lower complexity and constant modulus. Finally we also analyze the well known Walsh sequences which are also constant modulus.

### *A. Type I Sequences*

The expression for log-likelihood function involves pre-multiplication of the received signal vector with $[c_i(0)\ c_i(1)\ \ldots\ c_i(L_t-1)]$. Thus the outer $2L_g$ symbols only serve as separators from neighboring blocks. Further in the expression for log-likelihood function (8), the term corresponding to the main lobe of the raised cosine pulse, $R_{ii}(0)$ has no dependence on the separator symbols. We can therefore expect some performance improvement if we set all these outer terms to zero, keeping the total sequence energy constant. Further to get $N$ terms in the cross-correlation function, we split up the sequence into $N$ different parts, different from each other only by their phase. Thus the proposed GO sequence becomes,

$$\mathbf{s}_i = \begin{bmatrix}\mathbf{0} & \mathbf{x}^T & \mathbf{0} & \mathbf{x}^T\omega^{i-1}\ldots\mathbf{0} & \mathbf{x}^T\omega^{(i-1)(N-1)} & \mathbf{0}\end{bmatrix}^T \quad (11)$$

where

$$\mathbf{0} = [\underbrace{0\ 0\ \ldots\ 0}_{L_g \text{ times}}] \quad (12)$$

$$\text{and} \qquad \omega = \sqrt[N]{1} \quad (13)$$

and $\mathbf{x}$ is any sequence of length $L_x$ such that the normalized energy constraint is satisfied i.e., $\|\mathbf{s}_i\| = L_t + 2L_g$. It is easy to verify that the above sequences do satisfy the GOC. However since the length of $\mathbf{x}$ should be an integer, it is required that

$$\frac{L_t + 2L_g - (N+1)L_g}{N} = L_x \text{ is a positive integer} \quad (14)$$

Also note that for this sub-class,

$$R_{ii}(\tau) = N\sum_{n=1}^{L_x-\tau} x^*(n)x(n+\tau) \qquad \tau \geq 0 \quad (15)$$

which is independent of $i$. Henceforth we also drop $i$ from the subscript of $R(\tau)$. Note that $R(\tau)$ has conjugate symmetry property (i.e. $R(\tau) = R^*(-\tau)$).

In order to minimize the MCRB, we must minimize $\mathbb{E}\left[\frac{\partial^2}{\partial\epsilon^2}\Lambda(\epsilon)\right]$. Now,

$$\Lambda_{ij}(\epsilon) = \frac{1}{N_0}\left|\mathbf{c}_i^H\mathbf{r}_{oj}\right|^2 \quad (16)$$

$$= \frac{1}{N_0}\left|\sum_{n=0}^{L_t-1} c_i^*(n) r_{oj}((n+\epsilon)T)\right|^2 \quad (17)$$

$$= \frac{1}{N_0}\left|\sum_{n=0}^{L_t-1} c_i^*(n)\int_{-\infty}^{\infty} r_j(t)g(\epsilon T + nT - t)\mathrm{d}t\right|^2 \quad (18)$$

$$= \frac{1}{N_0}\sum_{m=0}^{L_t-1}\sum_{n=0}^{L_t-1} c_i^*(m)c_i(n)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{d}t_1\mathrm{d}t_2 \times r_j(t_1)r_j^*(t_2)g(mT+\epsilon T-t_1)g(nT+\epsilon T-t_2) \quad (19)$$

where $r_j(t)$ is the received signal given by,

$$r_j(t) = \sqrt{\frac{E_s}{N}}\sum_{k=1}^{N} h_{kj}\sum_{n=-L_g}^{L_t+L_g-1} c_k(n)g(t-nT-\epsilon T) + n_j(t) \quad (20)$$

Taking the second derivative with respect to $\epsilon$, using the above equation for $r_j(t)$ and taking the expectation with respect to the noise terms we get,

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial^2}{\partial\epsilon^2}\Lambda_{ij}(\epsilon)\right] \\
&= \frac{E_s}{NN_0}\sum_{m,n=0}^{L_t-1}\sum_{k,l=1}^{N}\sum_{m',n'=-L_g}^{L_t+L_g-1} h_{kj}^* h_{lj} c_i^*(m)c_i(n)c_k(m')c_l^*(n') \\
&\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t_1 - m'T - \epsilon T)g(t_2 - n'T - \epsilon T) \\
&\times \frac{\partial^2}{\partial\epsilon^2}\left(g(mT+\epsilon T - t_1)g(nT+\epsilon T - t_2)\right)\mathrm{d}t_1\mathrm{d}t_2 \\
&+ \frac{1}{N_0}\sum_{m=0}^{L_t-1}\sum_{n=0}^{L_t-1} c_i^*(m)c_i(n)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbb{E}[n_j(t_1)n_j^*(t_2)]\mathrm{d}t_1\mathrm{d}t_2 \\
&\times \frac{\partial^2}{\partial\epsilon^2}\left(g(mT+\epsilon T - t_1)g(nT+\epsilon T - t_2)\right) \qquad (21)
\end{aligned}$$

Since the noise $n_j(t)$ is white Gaussian process with variance $N_0$ and $\mathbb{E}[n_j^*(t_1)n_j(t_2)] = 0$ for $t_1 \neq t_2$. Now since $\dot{p}(t) = g(t)\otimes g'(t)$ and $\ddot{p}(t) = g(t)\otimes g''(t) = g'(t)\otimes g'(t)$ where the dots denote differentiation we get,

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial^2}{\partial\epsilon^2}\Lambda_{ij}(\epsilon)\right] \\
&= \frac{E_s T^2}{NN_0}\sum_{m,n=0}^{L_t-1}\sum_{k,l=1}^{N}\sum_{m',n'=-L_g}^{L_t+L_g-1} h_{kj}^* h_{lj} c_i^*(m)c_i(n)c_k(m')c_l^*(n') \\
&\quad\times [\ddot{p}((m-m')T)p((n-n')T) \\
&\qquad + p((m-m')T)\ddot{p}((n-n')T) \\
&\qquad + 2\dot{p}((m-m')T)\dot{p}((n-n')T)] \\
&\quad + 4\sum_{m=0}^{L_t-1}\sum_{n=0}^{L_t-1} c_i^*(m)c_i(n)\ddot{p}((m-n)T) \qquad (22)
\end{aligned}$$

where we have used the fact that the first and last $L_g$ symbols of the training sequences are zero (11). As $p(t)$ is assumed to be zero for $|t| > L_g$, $\dot{p}(t)$ is also zero for $|t| > L_g$. This is true for most ISI free pulses used in the literature. Using the change of variables as $m - m' = \tau$ and $n - n' = \kappa$ for the first term and $m - n = \mu$ in the second term and using the GOC property (7),

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial^2}{\partial\epsilon^2}\Lambda_{ij}(\epsilon)\right] &= \frac{2E_sT^2}{NN_0}|h_{ij}|^2\left[\sum_{\tau=-L_g}^{L_g} R(\tau)\ddot{p}(\tau T)\right. \\
&\quad \left. - \sum_{\tau,\kappa=-L_g}^{L_g} R(\kappa)R(\tau)\dot{p}(\tau T)\dot{p}(\kappa T)\right] \\
&\quad + 4\sum_{\mu=-L_g}^{L_g} R(\mu)\ddot{p}(\mu T) \qquad (23)
\end{aligned}$$

where we have used the fact that $R(\tau) = R^*(-\tau)$. Thus,

$$\mathbb{E}\left[\frac{\partial^2}{\partial\epsilon^2}\Lambda(\epsilon)\right] = (W+4N)\vartheta_2 - WNN_0\vartheta_1^2 \qquad (24)$$

where $W = 2T^2E_s/N_0 \sum\sum|h_{ij}|^2$, $\vartheta_1 = \mathbf{x}^H\mathbf{P}_1\mathbf{x}$ and $\vartheta_2 = \mathbf{x}^H\mathbf{P}_2\mathbf{x}$ where $\mathbf{P}_1$ and $\mathbf{P}_2$ are Toeplitz matrices given by,

$$\mathbf{P}_1 = \begin{bmatrix} \dot{p}(0) & \dots & \dot{p}(L_g) & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ \dot{p}(-L_g) & & \ddots & & & \\ 0 & \ddots & & & & \dot{p}(L_g) \\ \vdots & & & & & \vdots \\ 0 & \dots & & \dot{p}(-L_g) & \dots & \dot{p}(0) \end{bmatrix} \qquad (25)$$

and

$$\mathbf{P}_2 = \begin{bmatrix} \ddot{p}(0) & \dots & \ddot{p}(L_g) & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ \ddot{p}(-L_g) & & \ddots & & & \\ 0 & \ddots & & & & \ddot{p}(L_g) \\ \vdots & & & & & \vdots \\ 0 & \dots & & \ddot{p}(-L_g) & \dots & \ddot{p}(0) \end{bmatrix} \qquad (26)$$

Now since $\dot{p}(t)$ is an odd function while $\ddot{p}(t)$ is even, $\mathbf{P}_1^T = -\mathbf{P}_1$ and $\mathbf{P}_2^T = \mathbf{P}_2$ . This implies that $\vartheta_1$ is purely imaginary while $\vartheta_2$ is real. Denoting $\underline{\mathbf{x}} = [\Re(\mathbf{x}^T)\Im(\mathbf{x}^T)]^T$,

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial^2}{\partial\epsilon^2}\Lambda(\epsilon)\right] &= (W+4N)\underline{\mathbf{x}}^H\begin{bmatrix}\mathbf{P}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_2\end{bmatrix}\underline{\mathbf{x}} \\
&\quad + WNN_0\left(\underline{\mathbf{x}}^H\begin{bmatrix}\mathbf{0} & \mathbf{P}_1\\ -\mathbf{P}_1 & \mathbf{0}\end{bmatrix}\underline{\mathbf{x}}\right)^2 \qquad (27)
\end{aligned}$$

The second term is minimized when,

$$\underline{\mathbf{x}}^H\begin{bmatrix}\mathbf{0} & \mathbf{P}_1\\ -\mathbf{P}_1 & \mathbf{0}\end{bmatrix}\underline{\mathbf{x}} = 0 \qquad (28)$$

or equivalently $\Re(\mathbf{x}^H)\mathbf{P}_1\Im(\mathbf{x}) = 0$ which is always true thanks to the skew symmetric nature of $\mathbf{P}_1$. Also,

$$\vartheta_2 = \Re(\mathbf{x})^T\mathbf{P}_2\Re(\mathbf{x}) + \Im(\mathbf{x})^T\mathbf{P}_2\Im(\mathbf{x}) \qquad (29)$$

By the Rayleigh-Ritz theorem[5, Ch. 11], for any real symmetric matrix (like $\mathbf{P}_2$), the unit energy vector that minimizes $\mathbf{x}^T\mathbf{P}_2\mathbf{x}$ is the eigenvector corresponding to the smallest eigenvalue of $\mathbf{P}_2$. Thus the optimum values of both $\Re(\mathbf{x})$ and $\Im(\mathbf{x})$ subject to normalization constraints are given by,

$$\begin{aligned}
\Re(\mathbf{x}) &= \sqrt{\frac{(1-\beta^2)(L_t+2L_g)}{N}}\mathbf{u}_{\text{min}}(\mathbf{P}_2) \qquad (30)\\
\Im(\mathbf{x}) &= \sqrt{\frac{\beta^2(L_t+2L_g)}{N}}\mathbf{u}_{\text{min}}(\mathbf{P}_2)
\end{aligned}$$

where $\beta$ is any real number between 0 and 1 and we use $\mathbf{u}_{\text{min}}(.)$ operator to denote the eigenvector corresponding to the smallest eigenvalue. Since $\mathbf{P}_2$ is a Toeplitz matrix, it is easy to calculate the extreme eigenvalues using algorithms such as one described in [6]. Further, the minimum value of $\mathbf{x}^H\mathbf{P}_2\mathbf{x}$ is independent of the choice of $\beta$ (for example, we can simply assume $\beta = 1$).

Notice that, as shown in Fig. 1, the proposed GO sequences have a high peak to average power ratio (PAPR). This is mainly because we used a large number of zeros in $\mathbf{c}_i$ for

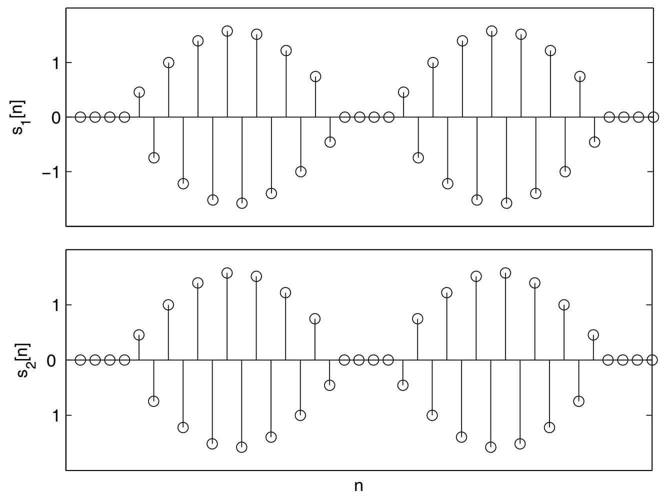


Fig. 1. Type I sequences: $\mathbf{s}_1$ and $\mathbf{s}_2$ for $N = 2$, $L_t = 32$ and $L_g = 4$.

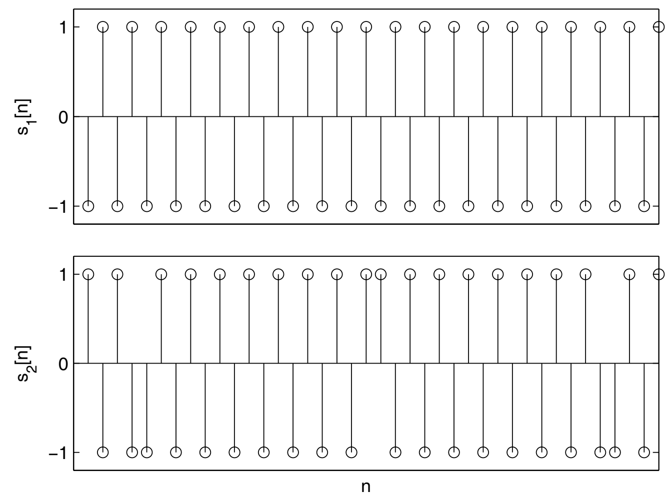


Fig. 2. Type II sequences: $\mathbf{s}_1$ and $\mathbf{s}_2$ for $N = 2$, $L_t = 32$ and $L_g = 4$.

separating it into $N$ different parts. This was required since we wanted these parts to not overlap when taking auto and cross correlation and thus sum up into $N$ terms which add up or cancel out respectively. Further the outer $2L_g$ symbols were set to zero because they did not contribute to $\Lambda(\epsilon_0')$.

Since high PAPR is usually an undesirable property, the next subsections propose two constant modulus sequences that can also significantly reduce the computational complexity.

### B. Type II Sequences

The problem of high PAPR can be solved if we do not insist on setting the outer symbols to zero. Thus like perfect sequences, we use the idea of cyclic prefix and suffix, do away with separating zeros and set all the symbols to be of unit magnitude. The Type II sequences thus look like

$$\mathbf{c}_i = \begin{bmatrix}\mathbf{x}^T & \mathbf{x}^T\omega^{i-1}\ldots\mathbf{x}^T\omega^{(i-1)(N-1)}\end{bmatrix}^T \tag{31}$$

and

$$\mathbf{s}_i = \begin{bmatrix}c_i(L_t - L_g) \ \ldots \ c_i(L_t - 1) \ \mathbf{c}_i^T \ c_i(0) \ \ldots \ c_i(L_g - 1)\end{bmatrix}^T \tag{32}$$

and each element of $\mathbf{x}$ has unit magnitude. It is easy to verify that the above sequences also satisfy the GOC. Similar to Type I sequences, the length of $\mathbf{x}$ imposes the following constraint,

$$\frac{L_t}{N} = L_x \text{ is a positive integer} \tag{33}$$

Note however that unlike Type I sequences, the autocorrelation function $R_{ii}(\tau)$ now becomes complex and depends on $i$. Thus the CRB expression (24) becomes a summation involving $h_{ij}$ and the optimal sequences can only be designed by averaging over all channel realizations. Proceeding as before, we get a similar optimization problem,

$$\min_{\mathbf{x}_i\in\{-1,1\}} \mathbf{x}^T\mathbf{P}_2\mathbf{x} + Q(\mathbf{x}) \tag{34}$$

where $Q(\mathbf{x}) = 4NN_0\sum_{\tau,\kappa}\sum_i \Im(R_{ii}(\tau))\Im(R_{ii}(\kappa))\dot{p}(\tau)\dot{p}(\kappa)$ is a quartic term resulting from the second term in (24). However since the second term depends only on the imaginary parts of $R_{ii}(\tau)$, it can be shown to be about $O(L_t)$ smaller than the first term (for $L_g \ll L_t$). Note that Equation (34) assumes, without loss of generality, that $\mathbf{x}$ is real (and therefore each element is either 1 or -1). As before, multiplication of $\mathbf{x}$ by a constant phase does not effect the optimization process.

The sequence design problem (34) belongs to the class of nonlinear integer programs which are usually NP hard to solve. If however, the quartic term is ignored the following observations lend a simple solution to the problem,

1) The sign of $\ddot{p}(i)$ alternates with $i$. This observation relates to the alternating convex and concave nature of the well known Nyquist pulses [7], [8].
2) Since $\mathbf{x}_i$ ($i$-th element of $\mathbf{x}$) is of unit magnitude and $\mathbf{P}_2$ is symmetric, we may reformulate the quadratic approximation as,

$$\min_{\mathbf{x}_i\in\{-1,1\}} \ddot{p}(1)\left(\sum_{i=1}^{N-1}\mathbf{x}_i\mathbf{x}_{i+1}\right) + \ddot{p}(2)\left(\sum_{i=1}^{N-2}\mathbf{x}_i\mathbf{x}_{i+2}\right) + \ldots \tag{35}$$

Since $\ddot{p}(1) > 0$, the minimum value of the first term is attained when $\mathbf{x}_i = -\mathbf{x}_{i+1}$. Similarly, the second term, which has a negative multiplicand $\ddot{p}(2)$, is minimized when $\mathbf{x}_i = \mathbf{x}_{i+2}$. Thus a sequence which satisfies $\mathbf{x}_i = (-1)^j\mathbf{x}_{i+j}$ for all $i$ and $j$ simultaneously minimizes all the terms in (35). It is easy to see that

$$\mathbf{x} = \begin{bmatrix}1 & -1 & 1 & -1 & \ldots & (-1)^{n-1}\end{bmatrix} \tag{36}$$

is the final approximate solution. Fig. 2 shows the overall type II sequence for $N = 2$, $L_t = 32$ and $L_g = 4$.

Interestingly, the estimator complexity is further reduced since the evaluation of $\mathbf{c}_i^H\mathbf{r}_{oj}(\epsilon)$ requires only $N$ multiplications. In fact, if we consider the number of multiplications as the only measure of estimator complexity, the DFT estimator used with Type II sequences requires only $MN^2Q$ multiplications as opposed to $MNQL_t$ multiplications required in [3].

We have thus been able to derive two near optimal GO sequences in $\mathbf{G}$. In fact, a first order expansion of the expression for MSE can be proved to be approximately equal to the asymptotic MCRB (see Appendix). However, note that although we can apply the interpolation technique of [2], we cannot apply the method of [3]. This is because unlike [3],

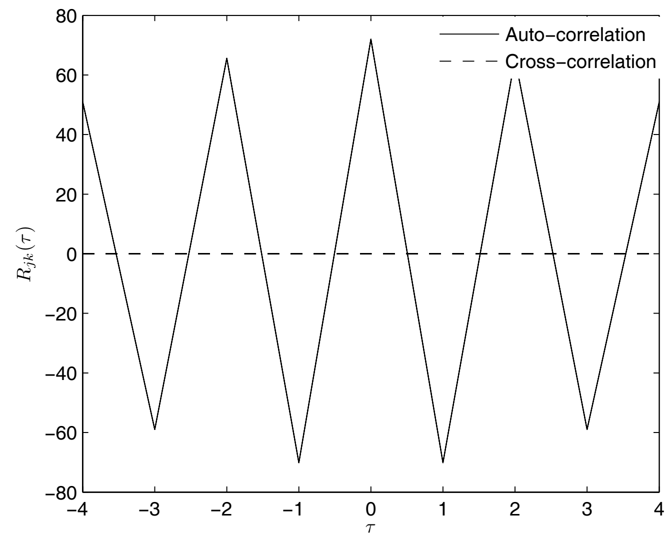


Fig. 3. Auto and Cross correlations of a set of length-64 Type I sequences. Note that all the auto correlation curves coincide and likewise for the cross correlation curves.

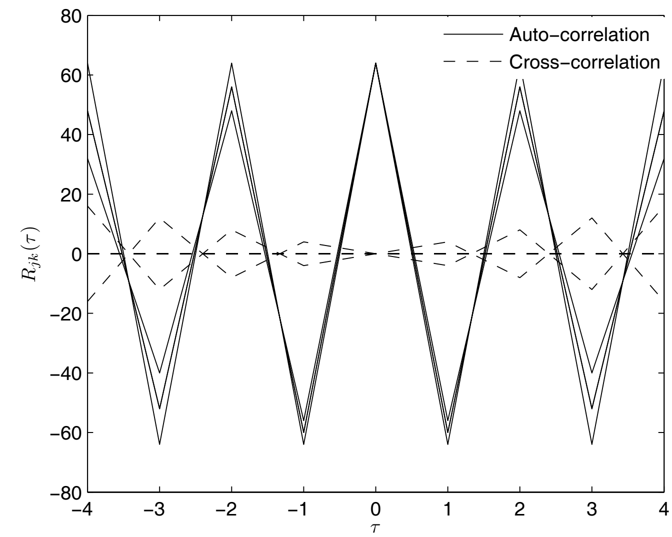


Fig. 4. Auto and Cross correlation of a set of length-64 Walsh sequences. Note that although the autocorrelation curves here do not overlap, most of the cross correlation curves do.

the ratio of log-likelihood functions at high SNRs for the two sampling instants given by,

$$f_2(\epsilon_0') = \frac{\Lambda(1)}{\Lambda(0)} = \frac{\left|\sum_{\tau=-L_g}^{L_g} R(\tau)p((\tau+1/2-\epsilon_0')T)\right|^2}{\left|\sum_{\tau=-L_g}^{L_g} R(\tau)p((\tau-\epsilon_0')T)\right|^2} \tag{37}$$

is not one-to-one in $\epsilon_0'$. Nevertheless, the estimator complexity changes only by a factor of two for considerable improvement in performance over [3], [2].

### *C. Walsh Sequences*

Another interesting class of constant modulus sequences that have good auto and cross-correlation properties are the Walsh sequences. Authors in [4] observe that the sign pattern of Walsh sequences resembles that of the optimal orthogonal sequences. In fact the performance of the Walsh sequences using ML estimators is also not far from optimal. This motivates us to consider Walsh sequences as candidates for low complexity timing estimation.

Typically, a Walsh sequence of length $2^n$ (for some $n > 2$) can have up to $2^n - 1$ sign changes. As suggested in [4], the $N$ Walsh sequences with the highest number of sign changes can be used as training sequences by adding cyclic suffixes and prefixes. For use in the low complexity estimator, it is insightful to compare their correlation properties with the Type I sequences (the Type II sequences cannot be used because $R_{ii}(\tau)$ is complex).

Fig. 3 and 4 show the auto and cross correlation of Type I and Walsh sequences respectively. We observe that although the Walsh sequences are not GO, their cross correlation values are small, especially for small shifts $\tau$. Thus Walsh sequence based training sequences, when plugged into the DFT based estimator will result in very little cross-channel interference. Further, their auto correlation function, though dependent on $i$, is also close to that of the (near optimal) Type I sequences. The MSE performance of Walsh sequences is therefore also expected to be only slightly worse than that of the Type I sequences.

## V. Simulation Results

The performance of the DFT estimator using the proposed GO sequences and the Walsh sequences has been compared with the estimator of [3] that uses perfect sequences by evaluating the estimation error using Monte Carlo simulations. The length of the guard band $L_g$ is assumed to be equal to $4$ for all three estimators. The MSE is approximated by averaging over $10^5$ estimates. We have taken $\epsilon$ to be uniformly distributed in $[-0.5, 0.5]$ and each $h_{ij}$ as independent complex Gaussian distributed random variable with zero mean and variance 0.5 each for the real as well as the imaginary parts. The pulse shape is assumed to be raised cosine with excess bandwidth $\alpha = 0.3$. The oversampling factor $Q$ is assumed to be $4$ for the proposed and DFT based methods and 2 for the [3]. The SNR is defined as $E_s/N_0$.

Fig. 5 shows the performance of estimator using the proposed Type I and Type II sequences and the Walsh sequences. The estimators of [3] and [2] and the MCRB are also plotted for comparison. We see that the DFT estimator utilizing the proposed GO sequences performs close to the MCRB. In fact the performance of Type I sequences is at least 5dB better than that of the perfect sequences. Further, the degradation in performance because of the constant modulus approximations (i.e., ignoring the quartic term in (34) for Type II sequences and suboptimal cross-correlation property of Walsh sequences) is also small while the performance is still much better than the perfect sequences.

## VI. Conclusion

Training sequences for data-aided symbol timing estimation in multi-input multi-output systems are designed. It is observed that the zero cross-correlation zone is a necessary condition for sequences which can be implemented at a low complexity. For convenience in minimizing the MCRB, we

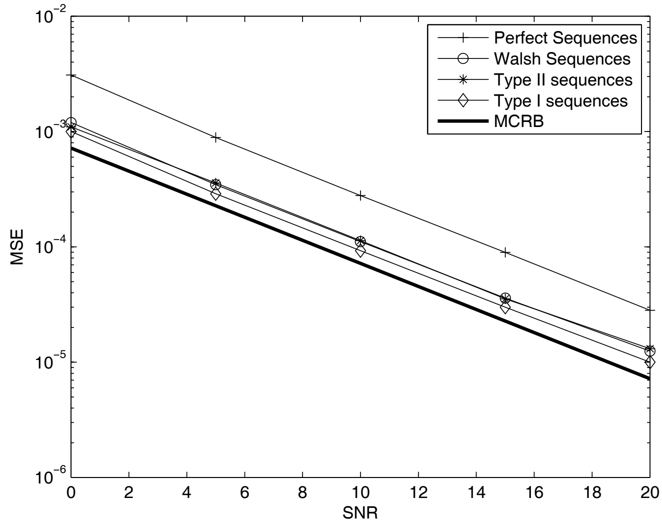


Fig. 5. Comparison of proposed, Walsh and perfect sequences for $N = 4$, $M = 2$, $L_t = 32$ and $L_g = 4$.

restrict ourselves to a subset of this class of sequences. Results show that the designed sequences perform better than the existing cyclic prefixes and suffixes based perfect sequences. Further, compared to the optimal orthogonal sequences, the proposed sequences have much less implementation complexity with only a marginal loss in performance. Two constant modulus sequences are also proposed which can further reduce the computational complexity with only a slight performance degradation.

## APPENDIX

Constructing the sequences $\mathbf{s}_i$ from (11) and (30) and plugging them back into (24), the asymptotic (high SNR) MCRB becomes,

$$MCRB_{DA} \approx -\frac{1}{2M\vartheta_2}\left(\frac{E_s}{N_0}\right)^{-1} \tag{39}$$

Now we show that the performance of the designed GO sequences is indeed close to the asymptotic MCRB given above. We derive a linearly approximated expression for MSE and show that it is indeed very close to the MCRB.

We begin by assuming high SNR so that the log likelihood function $\Lambda(\epsilon)$ is maximized at a value of $\epsilon$ close to the original value $\epsilon_0'$. This allows us to approximate $\Lambda(\epsilon)$ by its three term Taylor expansion,

$$\Lambda(\epsilon) = \Lambda(\epsilon_0') + \left.\frac{\mathrm{d}\Lambda(\epsilon)}{\mathrm{d}\epsilon}\right|_{\epsilon=\epsilon_0'}(\epsilon-\epsilon_0') + \frac{1}{2}\left.\frac{\mathrm{d}^2\Lambda(\epsilon)}{\mathrm{d}\epsilon^2}\right|_{\epsilon=\epsilon_0'}(\epsilon-\epsilon_0')^2 \tag{40}$$

In order to maximize $\Lambda(\epsilon)$, we set $\frac{\mathrm{d}\Lambda(\epsilon)}{\mathrm{d}\epsilon} = 0$ which gives the estimated value of $\epsilon_0'$ as,

$$\hat{\epsilon'}_0 = \epsilon_0' - \frac{\dot{\Lambda}(\epsilon_0')}{\mathbb{E}\left[\ddot{\Lambda}(\epsilon_0')\right]} \tag{41}$$

where the dots stand for derivative evaluated at $\epsilon_0'$ and we have approximated $\ddot{\Lambda}(\epsilon_0')$ by its noiseless version. The MSE is calculated as,

$$MSE \approx \mathbb{E}\left[(\hat{\epsilon_0'} - \epsilon_0')^2\right] = \frac{\mathbb{E}\left[\dot{\Lambda}(\epsilon_0')^2\right]}{\left(\mathbb{E}\left[\ddot{\Lambda}(\epsilon_0')\right]\right)^2} \tag{42}$$

The derivative of $\Lambda_{ij}(\epsilon)$ can be written as shown in (38) where,

$$s(t) = \sqrt{\frac{E_s}{N}}\sum_{k=1}^{N} h_{kj} \sum_{l=-L_g}^{L_t+L_g-1} c_k(l) g(t - lT - \epsilon_0' T) \tag{43}$$

Thus, squaring, substituting $\epsilon = \epsilon_0'$ and taking expectation with respect to noise terms, the simplified expression becomes

$$\mathbb{E}\left[\left(\dot{\Lambda}(\epsilon_0')\right)^2\right] = -\frac{2E_s}{N_0}\left(\sum_i \sum_j h_{ij}\right)\vartheta_2 \tag{44}$$

which has approximately the same magnitude as $\mathbb{E}\left[\ddot{\Lambda}(\epsilon_0')\right]$ evaluated in (24). Thus we see that the MSE turns out to be approximately same as the MCRB (i.e., the negative reciprocal of the expression in (24)) for sufficiently high SNRs.

## REFERENCES

[1] A. Naguib, V. Tarokh, N. Seshadri, and A. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1459-1478, Oct. 1998.
[2] Y.-C. Wu, S. Chan, and E. Serpedin, "Symbol-timing estimation in space-time coding systems based on orthogonal training sequences," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 603-613, Mar. 2005.
[3] K. Rajawat and A. Chaturvedi, "A low complexity symbol timing estimator for MIMO systems using two samples per symbol," *IEEE Commun. Lett.*, vol. 10, no. 7, pp. 525-527, July 2006.
[4] Y.-C. Wu and E. Serpedin, "Symbol timing estimation in MIMO correlated flat-fading channels," *Wireless Commun. Mobile Computing*, vol. 4, no. 7, pp. 773-790, Nov. 2004.
[5] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Prentice-Hall, 1980.
[6] A. Melman, "Extreme eigenvalues of real symmetric Toeplitz matrices," *Mathematics Computaion*, vol. 70, 2000.
[7] N. Beaulieu, C. Tan, and M. Damen, "A "better than" Nyquist pulse," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 367-368, Sept. 2001.
[8] A. Assalini and A. Tonello, "Improved Nyquist pulses," *IEEE Commun. Lett.*, vol. 8, no. 2, pp. 87-89, Feb. 2004.
[9] J. Riba, J. Sala, and G. Vazquez, "Conditional maximum likelihood timing recovery: estimators and bounds," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 835-850, Apr. 2001.

$$\begin{aligned}\frac{\mathrm{d}\Lambda_{ij}(\epsilon)}{\mathrm{d}\epsilon} &= \frac{1}{N_0}\sum_{m=0}^{L_t-1}\sum_{n=0}^{L_t-1} c_i^*(m)c_i(n)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_j(t_1)r_j^*(t_2)\frac{\mathrm{d}}{\mathrm{d}\epsilon}\left(g(mT+\epsilon T - t_1)g(nT+\epsilon T - t_2)\right)\mathrm{d}t_1\mathrm{d}t_2 \\ &= \frac{T}{N_0}\sum_{m=0}^{L_t-1}\sum_{n=0}^{L_t-1} c_i^*(m)c_i(n)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(s(t_1)+n_j(t_1)\right)\left(s^*(t_2)+n_j^*(t_2)\right)\mathrm{d}t_1\mathrm{d}t_2 \\ &\times\left[\dot{g}(mT+\epsilon T - t_1)g(nT+\epsilon T - t_2) + g(mT+\epsilon T - t_1)\dot{g}(nT+\epsilon T - t_2)\right]\end{aligned} \tag{38}$$

**Ketan Rajawat** (S'06) received his B.Tech.and M.Tech.degrees in Electrical Engineering from Indian Institute of Technology Kanpur in 2007. From 2006 to 2007, he was a teaching assistant and was working on symbol timing estimation in wireless communications at IIT Kanpur. Since August 2007, he has been working towards his Ph.D. degree at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His current research focuses on signal processing for wireless communications and network coding.

**Ajit K. Chaturvedi** (S'91-M'96-SM'03) received the B.Tech., M.Tech., and Ph.D. degrees in Electrical Engineering from Indian Institute of Technology Kanpur in 1986, 1988, and 1995, respectively. He served as a lecturer in the Department of Electronics Engineering at Institute of Technology, Banaras Hindu University, Varanasi from 1994 to 1996. Subsequently, he joined the faculty of the Department of Electronics and Computer Engineering at Indian Institute of Technology Roorkee. Since 1999, he has been teaching in the Department of Electrical Engineering at Indian Institute of Technology Kanpur. His research interests are in the area of Communications Theory and Wireless Communications. Dr. Chaturvedi has actively participated in the activities of the Uttar Pradesh section of IEEE. He has held several positions within the section and was its Chair during 2004-2005.