Carrier Frequency Estimation for Transmissions with Antenna Diversity

Young-Doo Kim†, Jae Kun Lim†, Chang-Ho Suh†, Eui-Rim Jeong†† and Yong H. Lee†*

† Division of Electrical Engineering, Korea Advanced Institute of Science and Technology
373-1 Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea

†† Hyundai Syscomm, Inc.
San 136-1, Ami-ri, Bubal-eub, Ichon-si, Kyoungki-do, 467-701, Republic of Korea

Abstract— **This paper deals with carrier frequency estimation for transmissions with antenna diversity. A joint maximum likelihood estimates (MLEs) of channel and frequency offset are derived and periodic space-time training sequences that can simplify the implementation of the ML frequency estimate, while minimizing the mean square error (MSE) of the estimate are designed. Statistical analysis indicates that the MLE is unbiased and almost achieves the Cramér-Rao bound (CRB). Simulation shows that MLE with optimal training sequences is preferable to one with arbitrary training sequences in mobile communications.**

I. Introduction

The stability requirements of recent communication systems on the carrier frequency offset between the oscillators of the transmitter and the receiver are stringent. It is required, for example, that the frequency offset between the carrier frequencies of a base station and a receiver is within 0.1 PPM in the 3rd generation partnership project (3GPP) recommendations for 3G wireless communication systems [1]. One way to relieve this need is to recover the carrier frequency at a receiver via digital signal processing

Various techniques have been proposed for carrier frequency recovery [2], [3]. Among these, data-aided techniques [4]–[10] using a training sequence (TS) are widely used because they can attain a good performance with short TSs. Most of the techniques proposed so far consider the transmission with a single antenna over such channels as additive white Gaussian noise (AWGN) [4]–[5], flat fading [6], [7], or frequency-selective fading channels [8]–[10]. In [10], it was observed that the estimators in [8]–[10] can be directly applied to frequency estimation of a system with transmitter antenna diversity when each transmission path experiences flat fading. When channels associated with multiple transmitter antennas are frequency-selective, these estimators should be properly modified. This paper deals with such modification. In particular, we derive joint ML channel and frequency offset estimators by extending the results in [9]. Two types of MLEs are developed: an MLE is derived first for an arbitrary TS, and then the result is modified for a periodic TS. The properties of these MLEs are analyzed, and a class of periodic TSs that can minimize the MSE of the MLE is classified.

This work was supported in part by the Korea Science and Engineering Foundation through the MICROS center at KAIST, Republic of Korea.

*The corresponding author.

II. Communication System Model

The baseband system considered in this paper is shown in Fig. 1. Here $p(t)$ is the baseband pulse shape, $c_i(t)$ is the channel impulse response, n(t) is AWGN, θ is the initial random phase and Δf represents the carrier frequency offset. The system has Γ transmitter antennas and one receiver antenna. It is assumed that a linear modulation (e.g. PSK or QAM) is employed. The output of the space-time (ST) coder at time k is given by

$$\mathbf{d}_{(k,N)} \triangleq \begin{bmatrix} \underline{d}_{(k,N)} \\ \underline{d}_{(k+1,N)} \\ \vdots \\ \underline{d}_{(k+N-1,N)} \end{bmatrix} \quad (1)$$

where $\underline{d}_{(k+j-1,N)} = [d^{(1)}_{k+j-1}, d^{(2)}_{k+j-1}, \cdots, d^{(\Gamma)}_{k+j-1}]$ and $d^{(i)}_k$ denotes the output of the i-th antenna at time k. The receiver filter output sampled at $t = kT_s$ (T_s denotes the symbol duration) is

$$r_k = e^{j(2\pi\Delta f k T_s+\theta)}\left(\sum_{l=0}^{L-1} g^{(1)}_{k,l} d^{(1)}_{k-l} + \cdots + \sum_{l=0}^{L-1} g^{(\Gamma)}_{k,l} d^{(\Gamma)}_{k-l}\right) + \eta_k \quad (2)$$

where $g^{(i)}_{k,l}$ is the impulse response of the equivalent channel from the i-th transmitter antenna to the receiver antenna at time k due to an impulse that is applied l time units earlier. It describes both $p(t)$ and $c_i(t)$ in the discrete time domain,

0-7803-7484-3/02/$17.00 ©2002 IEEE.

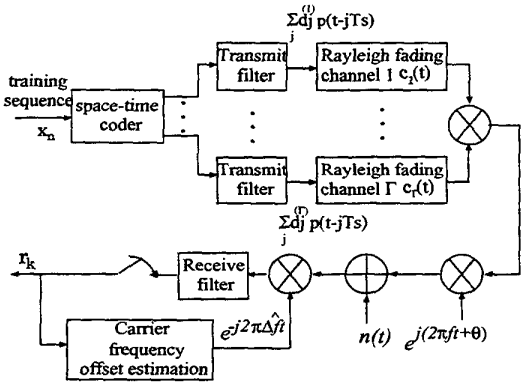

Fig. 1. Baseband system model.

and its duration is L. η_k is a zero-mean Gaussian noise with variance σ_η^2.

Suppose that a training sequence of length N, $\{x_n | \ 0 \leq n \leq N-1\}$ with $N \geq L\Gamma$, is inputted to the ST encoder. Then the encoder produces $\mathbf{d}_{(0,N)}$, which will be referred to as the *training* matrix. To compensate for the effect of multipath fading, it is necessary to start transmission at time $-L+1$: for $-L+1 \leq k \leq -1$, last $L-1$ columns of $\mathbf{d}_{(0,N)}$ are transmitted successively, i.e., $\underline{d}_{(k+N,N)}$ is sent at time k. Under the assumption that the channel is time-invariant over the training period, i.e., $g_{k,l}^{(i)} = g_l^{(i)}, -L+1 \leq k \leq N-1$, the received signal $\{r_0, r_1, \cdots, r_{N-1}\}$ can be written in vector form as

$$\mathbf{r} = \mathbf{\Theta}(\nu)\mathbf{D}_N\mathbf{g} + \boldsymbol{\eta} \tag{3}$$

where $\mathbf{r} = [r_0, r_1, \cdots, r_{N-1}]^T$, $\boldsymbol{\eta} = [\eta_0, \eta_1, \cdots, \eta_{N-1}]^T$, $\mathbf{\Theta}(\nu) = \mathrm{diag}[1, e^{j2\pi\nu}, \cdots, e^{j2\pi(N-1)\nu}]$ and $\mathbf{g}$ is an $L\Gamma$ dimensional vector defined as $\mathbf{g} = [g_0^{(1)}, g_0^{(2)}, \cdots, g_0^{(\Gamma)}, g_1^{(1)}, g_1^{(2)}, \cdots, g_{L-1}^{(\Gamma)}]^T$. $\nu = \Delta f T_s$ denotes the normalized frequency offset and $\mathbf{D}_N$ is an N-by-$L\Gamma$ dimensional matrix defined as $\mathbf{D}_N = [\mathbf{d}_{(0,N)}, \mathbf{d}_{(0,N)}^1 \cdots, \mathbf{d}_{(0,N)}^{L-1}]$, where $\mathbf{d}_{(0,N)}^i$ is the matrix formed by the i-th cyclic shift of the rows in $\mathbf{d}_{(0,N)}$, i.e., $\mathbf{d}_{(0,N)}^i = [\underline{d}_{(N-i,N)}^T, \cdots, \underline{d}_{(N-1,N)}^T, \underline{d}_{(0,N)}^T, \cdots, \underline{d}_{(N-i-1,N)}^T]^T$.

III. ML FREQUENCY OFFSET ESTIMATION

A. Derivation of Joint ML estimator

Given $\mathbf{g}$ and ν, the conditional joint probability density function Λ of $\mathbf{r}$ is given by

$$\Lambda(\mathbf{r}|\mathbf{g},\nu) = \frac{1}{(\pi\sigma_\eta^2)^N} \exp\left\{-\frac{1}{\sigma_\eta^2}||\mathbf{r} - \mathbf{\Theta}(\nu)\mathbf{D}_N\mathbf{g}||^2\right\} \tag{4}$$

and the joint ML estimation of $\mathbf{g}$ and ν can be found by maximizing the right-hand side of (4). This is equivalent to minimizing

$$||\mathbf{r} - \mathbf{\Theta}(\nu)\mathbf{D}_N\mathbf{g}||^2. \tag{5}$$

Suppose ν is fixed. Then the problem is reduced to the classical least squares minimization problem whose solution is, under the assumption that the columns of $\mathbf{D}_N$ are linearly independent,

$$\hat{\mathbf{g}}(\nu) = (\mathbf{D}_N^H\mathbf{D}_N)^{-1}\mathbf{D}_N^H\mathbf{\Theta}(\nu)^H\mathbf{r} \tag{6}$$

where H denotes the Hermitian transpose. Using (6) in (5), $||\mathbf{r} - \mathbf{\Theta}(\nu)\mathbf{D}_N\mathbf{g}||^2 = ||\mathbf{r} - \mathbf{\Theta}(\nu)\mathbf{B}\mathbf{\Theta}(\nu)^H\mathbf{r}||^2 = ||\mathbf{r}||^2 - ||\mathbf{\Theta}(\nu)\mathbf{B}\mathbf{\Theta}(\nu)^H\mathbf{r}||^2$ where

$$\mathbf{B} = \mathbf{D}_N(\mathbf{D}_N^H\mathbf{D}_N)^{-1}\mathbf{D}_N^H. \tag{7}$$

$\mathbf{B}$ is the orthogonal projection matrix onto the column space of $\mathbf{D}_N$. After removing the terms irrelevant to the estimation, minimizing (5) is equivalent to maximizing

$$h(\nu) = \mathbf{r}^H\mathbf{\Theta}(\nu)\mathbf{B}\mathbf{\Theta}(\nu)^H\mathbf{r}. \tag{8}$$

The ML estimator for ν is represented as

$$\hat{\nu} = \arg\max_\nu h(\nu). \tag{9}$$

We call this estimator the *MLE1*. This estimate can be obtained by exhaustively searching the frequency range $|v| \leq 0.5$, which is the acquisition range of the MLE1. Following [9], (9) is rewritten as

$$\hat{\nu} = \max_\nu \left\{-\rho(0) + 2\mathrm{Re}\left\{\sum_{m=0}^{\beta N-1} \rho(m)e^{-j2\pi m\nu}\right\}\right\} \tag{10}$$

where $\mathrm{Re}\{\cdot\}$ represents the real part and $\rho(m)$ is the weighted correlation of the received samples, defined as

$$\rho(m) = \begin{cases} \sum_{k=m+1}^{N} B(k-m,k)r_k r_{k-m}^*, & 0 \leq m \leq N-1 \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

In (11), $B(k,l)$ is the (k,l)th entry of an N-by-N matrix $\mathbf{B}$. The term in the bracket of (10) can be efficiently computed by using the fast Fourier transform (FFT) and computational load for evaluating MLE1 can be reduced. The estimator in (9) and (10) is a direct extension of the one in [9]: when $\Gamma = 1$, the former reduces to the latter.

For joint estimation of the channel and frequency offset, (10) should be evaluated prior to (6). Any TSs can be employed for the estimates, as long as $\mathbf{D}_N^H\mathbf{D}_N$ is nonsingular.

B. Mean and MSE of MLE1

Under the assumption of high signal-to-noise ratio (SNR), the mean and the MSE of MLE1 can be approximated as follows [3, pp. 343-344]:

$$\mathrm{E}[\hat{\nu}-\nu|\mathbf{g}] \approx -\frac{\mathrm{E}[h'(\nu)]}{\mathrm{E}[h''(\nu)]}. \tag{12}$$

$$\mathrm{E}[(\hat{\nu}-\nu)^2|\mathbf{g}] \approx -\frac{\mathrm{E}[(h'(\nu))^2]}{(\mathrm{E}[h''(\nu)])^2} \tag{13}$$

where h' and h'' denote the first and the second derivative of $h(\nu)$ in (8). By repeating almost verbatim the argument in [9]

$$\mathrm{E}[\hat{\nu}|\mathbf{g}] \approx \nu. \tag{14}$$

$$\mathrm{E}[(\hat{\nu}-\nu)^2|\mathbf{g}] \approx \frac{\sigma_\eta^2}{2\mathbf{y}^H(\mathbf{I}_N-\mathbf{B})\mathbf{y}} \tag{15}$$

where $\mathbf{y} = 2\pi\mathbf{M}\mathbf{D}_N\mathbf{g}$ with $\mathbf{M} = \mathrm{diag}[0,1,\ldots,N-1]$. (14) shows that the MLE1 is nearly unbiased. Under the assumption that the bias can be neglected, the CRB for the estimate in MLE1 can be derived again by repeating almost verbatim the argument in [9] and using the conditional density function in (4). The CRB (called *CRB1*) is given by

$$\mathrm{CRB1} = \frac{\sigma_\eta^2}{2\mathbf{y}^H(\mathbf{I}_N-\mathbf{B})\mathbf{y}}. \tag{16}$$

Since (15) coincides with (16), the MSE of MLE1 is close to the CRB.

IV. ML FREQUENCY OFFSET ESTIMATION FOR PERIODIC TSs

A. Derivation of MLE2

Suppose that a TS of length N consists of P identical subsequences of length K ($K \geq L\Gamma$): the TS is periodic and $N = KP$. In this case, the matrix $\mathbf{D}_N$ can be written as follows.

$$\mathbf{D}_N = [\mathbf{D}_K^T, \mathbf{D}_K^T, \cdots, \mathbf{D}_K^T]^T \tag{17}$$

where the K-by-$L\Gamma$ dimensional matrix $\mathbf{D}_K$ is given by

$$\mathbf{D}_K = [\mathbf{d}_{(0,K)}, \mathbf{d}_{(0,K)}^1, \cdots, \mathbf{d}_{(0,K)}^{L-1}]. \tag{18}$$

Here $\mathbf{d}_{(0,K)}$ and $\mathbf{d}_{(0,K)}^i$ are the matrices defined in (1) and (3), respectively. When $K = L\Gamma$, the matrix $\mathbf{B}$ in (7) is simplified as shown below.

Property 1. If $K = L\Gamma$, then the orthogonal projection matrix $\mathbf{B}$ in (7) becomes:

$$\mathbf{B} = \frac{1}{P}\begin{bmatrix} \mathbf{I}_K & \mathbf{I}_K & \cdots & \mathbf{I}_K \\ \mathbf{I}_K & \mathbf{I}_K & \cdots & \mathbf{I}_K \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_K & \mathbf{I}_K & \cdots & \mathbf{I}_K \end{bmatrix}. \tag{19}$$

TABLE I.
COMPUTATIONAL LOAD.

MLE1	real products	$2N(2N+2+\beta\mu\log_2\beta N)$
	real additions	$N(3N+1+3\beta\mu\log_2\beta N)$
MLE2	real products	$2N(N+K+\beta\mu\log_2(\beta N/K))/K$
	real additions	$N(2N+2K-2+3\beta\mu\log_2(\beta N/K))/K$

This is derived by using (17) in (7). When $\mathbf{B}$ is given by (19), MLE1 in (10) becomes

$$\hat{\nu} = \max_{\tilde{\nu}} \left\{ -\zeta(0) + 2\,\mathrm{Re}\left\{ \sum_{m=0}^{\beta P-1} \zeta(m) e^{-j2\pi m\tilde{\nu}} \right\} \right\} \tag{20}$$

where

$$\zeta(m) = \begin{cases} \frac{1}{P}\sum_{k=mK+1}^{N} r_k r_{k-mK}^*, & 0 \leq m \leq P-1, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

The acquisition range of this estimator, which will be referred to as *MLE2*, is $|v| \leq 1/(2K)$. Notice that P correlations $\{\zeta(m)\}$ are needed in (20), whereas $N(=KP)$ correlations $\{\rho(m)\}$ are in (10) and that FFT is performed over βP points in (20), whereas βN points in (10). Therefore, the computational complexity for implementing MLE1 is reduced approximately by a factor of $1/K$ (see Table I, $\mu = 1-(\log_2\beta + 2/\beta - 2)/\log 2(\beta N)$).

B. Mean and MSE of MLE2

MLE2 is unbiased like MLE1, because the result in (14) holds for any TSs for which $\mathbf{D}_N^H\mathbf{D}_N$ is nonsingular. The MSE of MLE2 can be computed from (15). After some calculation (see Appendix A), the MSE is expressed as follows.

$$\mathrm{E}[(\hat{\nu}-\nu)^2|\mathbf{g}] \approx \frac{3P\sigma_\eta^2}{2\pi^2N^2(P^2-1)\mathbf{g}^H\mathbf{D}_K^H\mathbf{D}_K\mathbf{g}}. \tag{22}$$

As in MLE1, (22) is identical to the CRB (called *CRB2*) of MLE2.

C. Training matrix Design

In this subsection, we classify a class of TS matrices that minimize the MSE $\mathrm{E}[(\hat{\nu}-\nu)^2]$, under the assumption that the channel $\mathbf{g}$ is complex Gaussian.

Property 2. The MSE of MLE2, $\mathrm{E}[(\hat{\nu}-\nu)^2]$, is minimized if $\mathbf{D}_K$ satisfies

$$\mathbf{D}_K^H\mathbf{D}_K = K\mathbf{I}_K. \tag{23}$$

The proof of this property is rather long and omitted due to page length limitation. Generation of a training matrix $\mathbf{d}_{(0,K)}$ satisfying (23) is addressed in the following properties.

multi path	L=1	L=2	L=4
	K=4, Γ=4	K=4, Γ=2	K=8, Γ=2
Training matrix $\mathbf{d}_{(0,K)}$	4x4 Hadamard matrix; or CAZAC sequence (length 4)	CAZAC sequence (length 4)	CAZAC sequence (length 8)

TABLE II.
TRAINING MATRICES THAT MINIMIZE THE MSE OF MLE2.

Property 3. When the channel is flat fading ($L = 1$), (23) is satisfied if and only if $1/\sqrt{K}\mathbf{d}_{(0,K)}$ is unitary.

This is a direct consequence of the fact that $\mathbf{D}_K = \mathbf{d}_{(0,K)}$ and $\mathbf{d}_{(0,K)}$ is a K-by-K square matrix when $L = 1$. For flat fading channels, well-known unitary matrices such as Hadamard matrix can be employed as $1/\sqrt{K}\mathbf{d}_{(0,K)}$, to yield the minimum MSE. In addition, Alamouti's 2-by-2 ST code [11] and *circulant* matrices whose first column is a *constant-amplitude zero autocorrelation* (CAZAC) sequence [3] of length K are also unitary; therefore, they can also be employed.

Property 4. When the channel is frequency-selective fading ($L \geq 2$), a training matrix $\mathbf{d}_{(0,K)}$ expressed as

$$\mathbf{d}_{(0,K)} = [\mathbf{c}_K^{(0)}, \mathbf{c}_K^{(L)}, \mathbf{c}_K^{(2L)}, \cdots, \mathbf{c}_K^{((\Gamma-1)L)}] \qquad (24)$$

satisfies (23), where $\mathbf{c}_K^{(0)}$ is a CAZAC sequence and $\mathbf{c}_K^{(i)}$ denotes i-th cyclic shift of the column vector $\mathbf{c}_K^{(0)}$.

Outline of the proof of this property is as follows. For $\mathbf{d}_{(0,K)}$ given by (24), each $\mathbf{c}_K^{(i)}$, $1 \leq i \leq K-1$, appears exactly once in the columns of $\mathbf{D}_K$ in (18). The orthogonality of the cyclic shifts of a CAZAC sequence implies that Property 2 is satisfied.

Table II shows some examples of the training matrix when $L = 1, 2, 4$.

V. APPLICATION TO THE FREQUENCY SELECTIVE FADING CHANNELS

The performances of MLE1 and MLE2 were assessed by applying them to a Rayleigh fading channel. For the simulation, the system model depicted in Fig. 2 was used. The number of training symbols N was 16, and the number of antenna Γ was 2. A raised-cosine filter with a rolloff of 0.5 was used for pulse shaping. For MLE1, a CAZAC sequence of length 16 was employed; the training matrix was

$$\mathbf{d}_{(0,16)}=[(1,1,1,1,j,-1,-j,1,-1,1,-1,1,-j,-1,j,1)^T \\ (-1,1,-1,1,-j,-1,j,1,1,1,1,1,j,-1,-j,1)^T].$$

For MLE2, $K = 8$ and the training matrix $\mathbf{d}_{(0,K)}$ was the one associated with $(K,\Gamma) = (8,2)$ in Table II; $\mathbf{d}_{(0,K)}$ successively transmitted twice. The channel response $g_{k,l}^{(i)}$ in (2), $i = 1, 2$, was obtained independently by extending the time-invariant channel in [9]. Specifically, $g_{k,l}^{(i)}$, $i = 1, 2$, is given by

$$g_{k,l}^{(i)} = \sum_{j=0}^{5} \xi_j(k) p(jT_s - \tau_j - t_0), -\infty \leq l \leq \infty \qquad (25)$$

where $\{\xi_j(k)\}$ and $\{\tau_j\}$ are the attenuation and delays of the paths, respectively, and t_0 is the timing phase, which was selected as equal to $\lfloor L/2 \rfloor T$—thereby guaranteeing that $\{g_{k,l}^{(i)} | 0 \leq l \leq L-1\}$ encompassed the L most significant channel elements. The normalized delays $\{\tau_j/T_s\}$ were set at { 0, 0.054, 0.135, 0.432, 0.621, 1.351}. For a given j, $\{\xi_j(k) | -\infty < k < \infty\}$ was a zero-mean complex Gaussian random process where the power spectral density (PSD) was bandlimited to a range $\pm f_D$, where f_D was the maximum Doppler shift. For different paths, $\{\xi_j(k) | 0 \leq j \leq 5\}$ were statistically independent and their variances were equal to { -3, 0, -2, -6, -8, -10 } (in decibels). $\{\xi_j(k)\}$ for the j-th path were generated by passing a complex Gaussian white noise through a baseband Doppler filter in [10]. After obtaining the time-varying channel $\{g_{k,l}^{(i)}\}$, $i = 1, 2$, MLE1 and MLE2 were implemented as follows.

For MLE1, L is upper bounded by N/Γ. L was set at 8 for this estimator, because $\{g_{k,l}^{(i)} | 0 \leq l \leq 7\}$ encompassed all significant channel elements, as observed in [9] for the corresponding time-invariant channel. For MLE2, $L \leq K/\Gamma$, and $L \leq 4$. Therefore, MLE2 should assume a channel with a duration of less than or equal to 4. For MLE1 and MLE2, the parameter β in (10) and (20) was fixed at 8. The normalized maximum Doppler shift $f_D T_s = 7.0 \times 10^{-3}$ in Fig. 3(a), and $f_D T_s = 1.5 \times 10^{-2}$

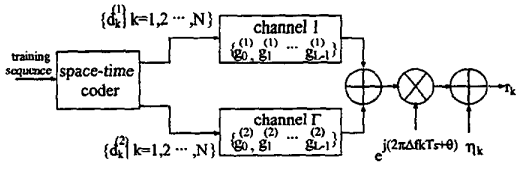

Fig. 2. System model used for simulation.

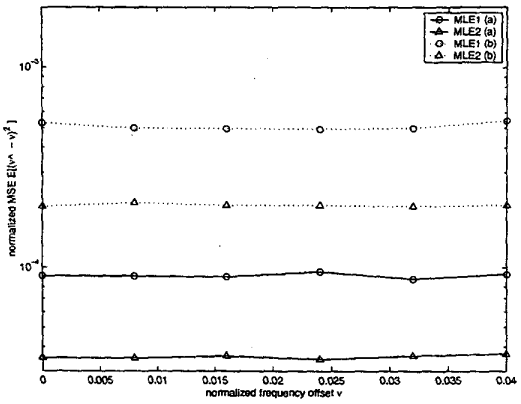

Fig. 3. Performance comparison for frequency-selective fading channel (E_b/N_0=15dB). (a) $f_D T_s = 7.0 \times 10^{-3}$ (b) $f_D T_s = 1.5 \times 10^{-2}$.

in Fig. 3(b), and E_b/N_0 is fixed to 15 dB. The normalized carrier frequency offset ν varies from 0 to 0.04. In the simulation, the MSE values were empirically estimated through 5,000 trials. The results are shown in Fig. 3. MLE2 that employs an optimal training matrix outperformed MLE1 whose training matrix was not optimally designed. The performances of the estimators were invariant in the range $|\nu| \leq 0.04$, which was the acquisition range of MLE2.

VI. CONCLUSION

Two types of joint ML channel and frequency offset estimates called MLE1 and MLE2, were derived for a system with transmitter diversity. MLE1 was developed for an arbitrary TS and MLE2 was derived for a periodic TS. The performance of these estimators were analyzed. In addition, for MLE2 a class of training matrices that minimize the MSE was classified. Simulation results indicated that MLE2 with an optimal training matrix can outperform MLE1. Derivation of optimal training matrices for MLE1 remains as further work.

APPENDIX

A. Derivation of the MSE of MLE2

(15) can be rewritten as

$$\mathrm{E}[(\hat{\nu}-\nu)^2|\mathbf{g}] \approx \frac{\sigma_\eta^2}{8\pi^2 \mathbf{g}^H \mathbf{D}_N^H \mathbf{M}(\mathbf{I}_N - \mathbf{B})\mathbf{M}\mathbf{D}_N \mathbf{g}}. \quad \text{(A1)}$$

Our objective is to show that $\mathbf{D}_N^H \mathbf{M}(\mathbf{I}_N - \mathbf{B})\mathbf{M}\mathbf{D}_N = N^2(P^2-1)/(12P)\mathbf{D}_K^H \mathbf{D}_K$. Define $\mathbf{M}_i = \mathrm{diag}[iK, iK+1, \cdots, iK+K-1]$, then $\mathbf{M} = \mathrm{diag}[\mathbf{M}_0, \mathbf{M}_1, \cdots, \mathbf{M}_{P-1}]$. After a direct calculation using the block structure of the matrices $\mathbf{M}(\mathbf{I}_N - \mathbf{B})\mathbf{M}$ becomes:

$$\frac{1}{\mathbf{P}} \begin{bmatrix} (\mathbf{P}-1)\mathbf{M}_0^2 & -\mathbf{M}_0\mathbf{M}_1 & \cdots & -\mathbf{M}_0\mathbf{M}_{P-1} \\ -\mathbf{M}_1\mathbf{M}_0 & (\mathbf{P}-1)\mathbf{M}_1^2 & \cdots & -\mathbf{M}_1\mathbf{M}_{P-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{M}_{P-1}\mathbf{M}_0 & -\mathbf{M}_{P-1}\mathbf{M}_1 & \cdots & (\mathbf{P}-1)\mathbf{M}_{P-1}^2 \end{bmatrix}. \quad \text{(A2)}$$

This implies that $\mathbf{M}(\mathbf{I}_N - \mathbf{B})\mathbf{M} = \mathbf{D}_K^H \mathbf{A} \mathbf{D}_K$ where $\mathbf{A} = \left(\sum_{k=0}^{P-1} \mathbf{M}_k^2 - \frac{1}{P}\sum_{k=0}^{P-1}\sum_{l=0}^{P-1} \mathbf{M}_k \mathbf{M}_l\right)$. Now, it is enough to show that $\mathbf{A} = N^2(P^2-1)/(12P)\mathbf{I}_K$. Since $\mathbf{A}$ is already diagonal, the mth diagonal term of $\mathbf{A}$ is

$$\sum_{k=0}^{P-1}(kK+m)^2 - \frac{1}{P}\sum_{k=0}^{P-1}\sum_{l=0}^{P-1}(kK+m)(lK+m) = N^2(P^2-1)/(12P). \quad \text{(A3)}$$

REFERENCES

[1] 3G TS 25.211 (V3.5.0), Physical channels and mapping of transport channels onto physical channels (FDD), http://www.3gpp.org/ftp/Specs/2000-12/R1999/25_series/25211-350.zip, Dec. 2000.

[2] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, New York, 1997.

[3] H. Meyr, M. Moeneclaey, ans S. A. Fechtel, *Digital Communication Receivers*, John Wiley & Sons, New York, 1998.

[4] M. Luise and R. Reggiannini, "Carrier frequency recovery in all-digital modems for burst-mode transmissions,"" *IEEE Trans. Commun.*, vol. 43, pp. 1169–1178, Feb./Mar./Apr. 1995.

[5] U. Mengali and M. Morelli, "Data-aided frequency estimation for burst digital transmission,"" *IEEE Trans. Commun.*, vol. 45, pp. 23–25, Jan. 1997.

[6] W. Y. Kuo and M. P. Fitz, "Frequency offset compensation of pilot symbol assisted modulation in frequency flat fading," *IEEE Trans. Commun.*, vol. 45, pp. 1412–1417, Nov. 1997.

[7] M. Morelli, U. Mengali, and G. M. Vitetta, "Further results in carrier frequency estimation for transmissions over flat fading channels," *IEEE Commun. Letters*, vol. 2, pp. 327-330, Dec. 1998.

[8] M. G. Hebley and D. P. Taylor, "The effect of diversity on a burst-mode carrier frequency estimator in the frequency-selective multipath channels," *IEEE Trans. Commun.*, vol. 46, pp. 553–560, Apr. 1998.

[9] M. Morelli and U. Mengali, "Carrier frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, pp. 1580–1589, Sept. 2000.

[10] E. R. Jeong, S. K. Jo, and Y. H. Lee, "Least squares frequency estimation in frequency-selective channels and its applications to transmissions with antenna diversity," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2369–2380, Dec. 2001.

[11] S. M. Alamouti, "A simple transmitter diversity scheme for wierless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, 1998.

[12] W. Fleming, *Functions of Several Variables*, Second Ed., Springer-Verlag, New York, 1977.