DOUBLY-SELECTIVE MULTIUSER CHANNEL ESTIMATION USING SUPERIMPOSED TRAINING AND DISCRETE PROLATE SPHEROIDAL BASIS EXPANSION MODELS

Shuangchi He   Jitendra K. Tugnait

Department of Electrical & Computer Engineering, Auburn University, Auburn, AL 36849, USA

ABSTRACT

Channel estimation for multiuser doubly-selective channels is considered using superimposed training. The time-varying channel is assumed to be described by a discrete prolate spheroidal basis expansion model (DPS-BEM). A user-specific periodic training sequence is arithmetically added (superimposed) at a low power to each user's information sequence at the transmitter before modulation and transmission. A two step approach is adopted where in the first step we estimate the channel using only the first-order statistics of the observations. Using the estimated channel from the first step, a Viterbi detector is used to estimate the information sequence. In the second step a deterministic maximum likelihood (DML) approach is used to iteratively estimate the multiuser channel and the information sequences sequentially.

Index Terms— Doubly-selective channels, channel estimation, basis expansion models, multiuser channels

1. INTRODUCTION

Consider a doubly-selective (time- and frequency-selective) multiuser finite impulse response (FIR) linear channel with K inputs (users) and N outputs, resulting in a multiple-input multiple-output (MIMO) formulation. Let $\{s_k(n)\}$ denote the k-th user's transmitted symbols which is input to the MIMO channel with the k-th user's discrete-time impulse response $\{\mathbf{h}_k(n;l)\}$. Then the symbol-rate, N-column channel output vector (resulting from N receive antennas, e.g.) is given by

$$\mathbf{x}(n) = \sum_{k=1}^{K} \sum_{l=0}^{L} \mathbf{h}_k(n;l) s_k(n-l). \tag{1}$$

A parsimonious representation of time-varying channels is provided by basis expansion models (BEM) where one assumes

$$\mathbf{h}_k\left(n;l\right) = \sum_{q=1}^{Q} \mathbf{h}_{qk}\left(l\right) u_q\left(n\right) \tag{2}$$

where $u_q(\cdot)$ is the q-th basis function, and $\mathbf{h}_{qk}(l)$'s are fixed over the data block. In the complex exponential basis expansion model (CE-BEM) [4], for a data block length of T symbols with symbol interval T_s sec., one chooses $u_q\left(n\right) = e^{j\omega_q n}, \omega_q := 2\pi\left[q - \frac{Q+1}{2}\right]/T$, $L := \lfloor \tau_d/T_s \rfloor$, and $Q := 2\lceil f_d T T_s \rceil + 1$ when the underlying continuous-time channel has a delay spread of τ_d sec. and Doppler spread of f_d Hz. In discrete prolate spheroidal BEM (DPS-BEM), the i-th DPS vector $\boldsymbol{u}_i := \left[u_i\left(0\right), \cdots, u_i\left(T-1\right)\right]^T$ (called Slepian sequence in [9], which is time-windowed DPS sequence) is the i-th eigenvector of a matrix $\mathbf{C}$ [8]: $\mathbf{C}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$, where $[\mathbf{C}]_{n,m} = \frac{\sin[2\pi(n-m)f_d T_s]}{[\pi(n-m)]}$ is the (n,m)-th entry of $\mathbf{C}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_T$ are the eigenvalues of $\mathbf{C}$. The DPS sequences $u_q(n)$ are orthonormal over the finite time interval $[0, T-1]$.

This work was supported by NSF under Grant ECS-0424145.

The rectangular window of the truncated DFT-based CE-BEM model introduces spectral leakage [7], [9]. The energy at each individual frequency leaks to the full frequency range, resulting in significant amplitude and phase distortion at the beginning and the end of the data block [9]. DPS sequences are a good alternative as a basis set to approximate bandlimited channels alleviating the spectral leakage of CE-BEM [9]. In this case one takes $Q = \lceil 2 f_d T_s T \rceil + 1$ [9].

The noisy measurement of $\mathbf{x}(n)$ is given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n). \tag{3}$$

Our objective is to recover $\{s_k\left(n\right)\}$ given noisy $\{\mathbf{y}\left(n\right)\}$. In several approaches this requires knowledge of the channel impulse response. Recently superimposed training based approaches have been investigated where for the k-th user, one takes

$$s_k\left(n\right) = b_k\left(n\right) + c_k\left(n\right), \tag{4}$$

where $\{c_k\left(n\right)\}$ is a training (pilot) sequence added (superimposed) at a low power to the information sequence $\{b_k\left(n\right)\}$ at the transmitter before modulation and transmission. There is no loss in data transmission rate, unlike the conventional time-multiplexed training. Time-invariant multiuser channel estimation using superimposed training can be found in [5] and [1]. CE-BEM-based doubly-selective multiuser channel estimation using superimposed training can be found in [6]. DPS-BEM-based single-user channel estimation has recently been reported in [2]. Ref. [9] is the first to use DPS-BEM for doubly-selective channel estimation using time-multiplexed training.

Objectives and Contributions: We first extend the first-order statistics-based approach of [2] to multiuser systems, where the information sequences from all users act as interference in channel estimation. We then further extend it to an iterative deterministic maximum likelihood (DML) approach. In subsequent iterations, the estimate of the information sequences are exploited to enhance channel estimation, and thus information sequence detection. The DML approach has been used before in [5] and [6] also for time-invariant and CE-BEM-based doubly-selective multiuser channels, respectively.

Notation: Superscripts H, $*$ and T denote the complex conjugate transpose, complex conjugation, and transpose operations, respectively. $\delta(\cdot)$ is the Kronecker delta function and $\mathbf{I}_N$ is the $N \times N$ identity matrix. The symbol $\otimes$ denotes the Kronecker product.

2. FIRST ORDER STATISTICS-BASED ESTIMATOR

In this section we extend the first-order statistics-based approach using DPS-BEM of [2] to doubly-selective multiuser systems. The main idea is to pick user-specific training sequences so that the problem of channel estimation is decoupled across various users — this allows us to use the single user superimposed training-based approach outlined in [2]. Our approach is to assign distinct cycle frequencies of the periodic training sequences to distinct users. Suppose that for every user k, $\{c_k(n)\}$ is periodic with period $P = \tilde{P}K$ where $\tilde{P}$ is a positive integer. Then, in general

$$c_k(n) = \sum_{m'=0}^{P-1} c_{m'k} e^{j(2\pi m'/P)n} \quad \forall n$$

where $c_{m'k} := P^{-1}\sum_{n=0}^{P-1} c_k(n)e^{-j(2\pi m'/P)n}$. Pick $\{c_k(n)\}$ so that only $\tilde{P}$ coefficients (out of total P) $c_{m'k}$, associated with $\tilde{P}$ distinct frequencies, are nonzeros. For instance, we may choose

$$c_k(n) = \sum_{m=0}^{\tilde{P}-1} c'_{mk} e^{j\alpha_{mk}n}, \quad \alpha_{mk} := 2\pi(Km + k - 1)/P, \tag{5}$$

for suitably chosen $c'_{mk} \neq 0 \ \forall m, k$. One specific choice may be found in [5]. [One may take $c'_{mk} = c'_{m1}$ for $k \geq 2$ and $\forall m$.]

We now state our model assumptions:

- **(H1)** The time-varying channel $\{\mathbf{h}_k(n;l)\}$ satisfies (2) using the DPS-BEM representation. Also $N \geq 1$.
- **(H2)** The information sequences $\{b_k(n)\}$ are zero-mean, finite-alphabet, i.i.d. with $E\{|\, b_k(n)|^2\} = \sigma_{bk}^2$ and mutually independent for $k = 1, 2, \cdots, K$.
- **(H3)** The measurement noise $\{\mathbf{v}(n)\}$ in (3) is possibly **nonzero-mean** ($E\{\mathbf{v}(n)\} = \mathbf{m}$), white complex Gaussian, uncorrelated with $\{b_k(n)\}$, with $E\{[\mathbf{v}(n+\tau) - \mathbf{m}][\mathbf{v}(n) - \mathbf{m}]^H\} = \sigma_v^2 \mathbf{I}_N \delta(\tau)$. The mean vector $\mathbf{m}$ is unknown.
- **(H4)** The superimposed training sequences $c_k(n) = c_k(n + mP)$ $\forall m, n$ are non-random periodic sequences with period P and average power $\sigma_{ck}^2 := \sum_{n=0}^{P-1} |c_k(n)|^2 / P$, satisfying (5) such that $c'_{mk} \neq 0 \, \forall m, k$, and $\tilde{P}$ is an integer with $P = \tilde{P}K$.

Using (1)-(5) we have (here channel is non-random)

$$E\{\mathbf{y}(n)\} = \sum_{k=1}^{K}\sum_{m=0}^{\tilde{P}-1}\sum_{q=1}^{Q} \underbrace{\left[\sum_{l=0}^{L} \mathbf{h}_{qk}(l) c_{mk} e^{-j\alpha_{mk}l}\right]}_{=:\mathbf{d}_{mqk}} \times u_q(n)\, e^{j\alpha_{mk}n} + \mathbf{m}, \quad \forall n. \tag{6}$$

It then follows that

$$\mathbf{y}(n) = \sum_{k=1}^{K}\sum_{m=0}^{\tilde{P}-1}\sum_{q=1}^{Q} \mathbf{d}_{mqk} u_q(n)\, e^{j\alpha_{mk}n} + \mathbf{m} + \mathbf{e}(n)$$

where $\{\mathbf{e}(n)\}$ is a zero-mean random sequence. Define the cost function $J = \sum_{n=0}^{T-1} \|\mathbf{e}(n)\|^2$. Choose $\mathbf{d}_{mqk}$'s and $\mathbf{m}$ to minimize J. Then we must have

$$\left.\frac{\partial J}{\partial \mathbf{d}^*_{mqk}}\right|_{\mathbf{d}_{mqk} = \hat{\mathbf{d}}_{mqk}} = 0, \quad \left.\frac{\partial J}{\partial \mathbf{m}^*}\right|_{\mathbf{m} = \hat{\mathbf{m}}} = 0, \tag{7}$$

which leads to

$$\sum_{k'=1}^{K}\sum_{q'=1}^{Q}\sum_{m'=0}^{\tilde{P}-1} \hat{\mathbf{d}}_{m'q'k'} \left[\sum_{n=0}^{T-1} u_{q'}(n)\, u_q(n)\, e^{j\left(\alpha_{m'k'} - \alpha_{mk}\right)n}\right] + \hat{\mathbf{m}} \sum_{n=0}^{T-1} u_q(n)\, e^{-j\alpha_{mk}n} = \sum_{n=0}^{T-1} \mathbf{y}(n)\, u_q(n)\, e^{-j\alpha_{mk}n}. \tag{8}$$

The time-limited DPS sequences are windowed (using rectangular windows) versions of infinite DPS sequences that are exactly band-limited to $[-f_d T_s, f_d T_s]$ [8],[9]. Therefore, taking the time-limited DPS sequences as approximately band-limited to $[-f_d T_s, f_d T_s]$, for $f_d T_s \ll 1/P$ and T either a multiple of P or "large", we have

$$\sum_{n=0}^{T-1} u_q(n)\, e^{-j\alpha_{mk}n} \approx 0 \quad \forall \alpha_{mk} \neq 0, \tag{9}$$

$$\sum_{n=0}^{T-1} u_{q'}(n)\, u_q(n)\, e^{j\left(\alpha_{m'k'} - \alpha_{mk}\right)n} \approx \delta(k' - k)\, \delta(m' - m)\, \delta(q' - q). \tag{10}$$

Since $\alpha_{mk} = 0$ only happens when $m = 0$ and $k = 1$, then for $\alpha_{mk} \neq 0$, using (8)-(10), an estimate $\hat{\mathbf{d}}_{mqk}$ of $\mathbf{d}_{mqk}$, follows as [2]

$$\hat{\mathbf{d}}_{mqk} = \sum_{n=0}^{T-1} \mathbf{y}(n) u_q(n)\, e^{-j\alpha_{mk}n} \tag{11}$$

where omission of $\alpha_{mk} = 0$ allows us to decouple estimation of $\mathbf{d}_{mqk}$'s and $\mathbf{m}$. It is easily seen from (6) and (11) that $E\{\hat{\mathbf{d}}_{mqk}\} = \mathbf{d}_{mqk}$. Since $\mathbf{m}$ is unknown, we omit the term $m = 0$ from further discussion.

For $1 \leq m \leq \tilde{P} - 1$, $1 \leq k \leq K$, and $0 \leq l \leq L$, we define

$$\mathbf{D}_{mk} := \begin{bmatrix} \mathbf{d}^T_{m1k} & \cdots & \mathbf{d}^T_{mQk} \end{bmatrix}^T, \tag{12}$$

$$\mathbf{H}_{kl} := \begin{bmatrix} \mathbf{h}^H_{1k}(l) & \cdots & \mathbf{h}^H_{Qk}(l) \end{bmatrix}^H,$$

$$\mathbf{V}_k := \begin{bmatrix} 1 & e^{-j\alpha_{1k}} & \cdots & e^{-j\alpha_{1k}L} \\ 1 & e^{-j\alpha_{2k}} & \cdots & e^{-j\alpha_{2k}L} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\alpha_{(\tilde{P}-1)k}} & \cdots & e^{-j\alpha_{(\tilde{P}-1)k}L} \end{bmatrix},$$

$$\mathcal{C}_k := \left(\mathrm{diag}\left\{c'_{1k}, \cdots, c'_{(\tilde{P}-1)k}\right\} \mathbf{V}_k\right) \otimes \mathbf{I}_{NQ},$$

$$\mathcal{H}_k = \begin{bmatrix} \mathbf{H}^H_{k0} & \cdots & \mathbf{H}^H_{kL} \end{bmatrix}^H,$$

$$\mathcal{D}_k = \begin{bmatrix} \mathbf{D}^H_{1k} & \cdots & \mathbf{D}^H_{(\tilde{P}-1)k} \end{bmatrix}^H. \tag{13}$$

By the definition of $\mathbf{d}_{mqk}$ in (6), we have

$$\mathcal{C}_k \mathcal{H}_k = \mathcal{D}_k.$$

Since α_{mk}'s are distinct and $c'_{mk} \neq 0 \, \forall m, k, (m \neq 0)$, rank $(\mathcal{C}_k) = NQ(L + 1)$ if $\tilde{P} \geq L + 2$; hence, we can determine $\mathbf{h}_{qk}(l)$'s uniquely. Define $\hat{\mathbf{D}}_{mk}$ as in (12) with $\mathbf{d}_{mqk}$ replaced with $\hat{\mathbf{d}}_{mqk}$ and define $\hat{\mathcal{D}}_k$ as in (13) with $\mathbf{D}_{mk}$ replaced with $\hat{\mathbf{D}}_{mk}$. Then the estimate of $\mathcal{H}_k$ is given by

$$\hat{\mathcal{H}}_k = (\mathcal{C}_k^H \mathcal{C}_k)^{-1} \mathcal{C}_k^H \hat{\mathcal{D}}_k. \tag{14}$$

Denote the corresponding estimate of $\mathbf{h}_{qk}(l)$ as $\hat{\mathbf{h}}_{qk}(l)$. Following (2), the estimate of the time-varying channel is given by

$$\hat{\mathbf{h}}_k(n; l) = \sum_{q=1}^{Q} \hat{\mathbf{h}}_{qk}(l) u_q(n).$$

Performance Analysis: We assume that the channel satisfies the following assumption:

(H5) The time-varying channels $\{\mathbf{h}_k(n;l)\}$ are zero-mean, complex Gaussian with $E\{\mathbf{h}_k(n;l)\,\mathbf{h}_k^H(n;l)\} = \sigma_h^2 \mathbf{I}_N$, and mutually independent for distinct l's and different users.

If the true channel follows (2), the mean square error (MSE) in channel estimation for user k can be shown to be given by

$$\begin{aligned}\mathrm{MSE}_k &:= \frac{1}{T}\sum_{n=0}^{T-1}\sum_{l=0}^{L} E\left\{\left\|\mathbf{h}_k(n;l)-\hat{\mathbf{h}}_k(n;l)\right\|^2\right\}\\ &= \left[(L+1)\sum_{k=1}^{K}\sigma_{hk}^2\sigma_{bk}^2+\sigma_v^2\right]NQT^{-1}\\ &\quad\times \mathrm{tr}\left\{\mathbf{V}_k^H\left(\mathrm{diag}\left\{\left|c'_{1k}\right|^2,\cdots,\left|c'_{(\tilde{P}-1)k}\right|^2\right\}\mathbf{V}_k\right)^{-1}\right\}.\end{aligned} \tag{15}$$

3. ITERATIVE DETERMINISTIC MAXIMUM LIKELIHOOD (DML) APPROACH

Since the training and information sequences pass through an identical channel, this fact can be exploited to enhance the channel estimation performance. We now consider joint channel and information sequence estimation performance via an iterative DML approach assuming that the noise $\mathbf{v}(n)$ is complex Gaussian.

We define the following

$$\mathbf{Y} := \begin{bmatrix}\mathbf{y}^T(T-1) & \cdots & \mathbf{y}^T(L)\end{bmatrix}^T$$

$$\mathbf{s} = [s_1(T-1),\cdots,s_K(T-1),s_1(T-2),\cdots s_K(0)]^T.$$

$$\Sigma_n := \begin{bmatrix} u_1(n)\,\mathbf{I}_N & \cdots & u_Q(n)\,\mathbf{I}_N\end{bmatrix},$$

$$\mathcal{T}(\mathbf{s}) := \begin{bmatrix} s_1(T-1)\Sigma_{T-1} & \cdots & s_1(T-L-1)\Sigma_{T-1} & \cdots & s_K(T-L-1)\Sigma_{T-1}\\ s_1(T-2)\Sigma_{T-2} & \cdots & s_1(T-L-2)\Sigma_{T-2} & \cdots & s_K(T-L-2)\Sigma_{T-2}\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ s_1(L)\Sigma_L & \cdots & s_1(0)\Sigma_L & \cdots & s_K(0)\Sigma_L\end{bmatrix}$$

$$\mathcal{H} = \begin{bmatrix}\mathcal{H}_1^T & \cdots & \mathcal{H}_K^T\end{bmatrix}^T,$$

$$\tilde{\mathbf{V}} = \begin{bmatrix}\tilde{\mathbf{v}}^T(T-1) & \cdots & \tilde{\mathbf{v}}^T(L)\end{bmatrix}^T$$

where $\tilde{\mathbf{v}}(n) := \mathbf{v}(n) - \mathbf{m}$. By (1)-(3), we then have the linear model

$$\mathbf{Y} = \mathcal{T}(\mathbf{s})\mathcal{H} + \tilde{\mathbf{V}} + \mathcal{M} \tag{16}$$

where $\mathcal{M} := \begin{bmatrix}\mathbf{m}^T,\cdots,\mathbf{m}^T\end{bmatrix}^T$. If we further define

$$\mathcal{F}(\mathcal{H}) := \begin{bmatrix}\mathbf{h}_1(T-1;0) & \cdots & \mathbf{h}_K(T-1;0) & \cdots & \mathbf{h}_K(T-1;L) & & \\ & & \ddots & \ddots & & \ddots & \\ & & & \mathbf{h}_1(L;0) & \cdots & \mathbf{h}_1(L;L) & \cdots & \mathbf{h}_K(L;L)\end{bmatrix}$$

we obtain another linear model as

$$\mathbf{Y} = \mathcal{F}(\mathcal{H})\mathbf{s} + \tilde{\mathbf{V}} + \mathcal{M}. \tag{17}$$

We consider the joint estimation

$$\left\{\hat{\mathcal{H}},\hat{\mathbf{s}},\hat{\mathbf{m}}\right\} = \arg\min_{\mathcal{H},\mathbf{s},\mathbf{m}}\|\mathbf{Y}-\mathcal{T}(\mathbf{s})\,\mathcal{H}-\mathcal{M}\|^2. \tag{18}$$

Under a white Gaussian noise assumption, the nonlinear least-squares optimization (18) yields the DML parameter estimator. Using (16) and (17), we have a separable nonlinear least-squares problem that can be solved sequentially as follows. At iteration j, with an initial guess of the channel $\mathcal{H}^{(j)}$, and the mean $\mathbf{m}^{(j)}$, the algorithm estimates the input sequence $\mathbf{s}^{(j)}$ and the channel $\mathcal{H}^{(j+1)}$ and mean $\mathbf{m}^{(j+1)}$ for the next iteration by

$$\mathbf{s}^{(j)} = \arg\min_{\mathbf{s}\in\mathcal{S}}\left\|\mathbf{Y}-\mathcal{F}\left(\mathcal{H}^{(j)}\right)\mathbf{s}-\mathcal{M}^{(j)}\right\|^2, \tag{19}$$

$$\mathcal{H}^{(j+1)} = \arg\min_{\mathcal{H}}\left\|\mathbf{Y}-\mathcal{T}\left(\mathbf{s}^{(j)}\right)\mathcal{H}-\mathcal{M}^{(j)}\right\|^2, \tag{20}$$

$$\mathbf{m}^{(j+1)} = \arg\min_{\mathbf{m}}\left\|\mathbf{Y}-\mathcal{T}\left(\mathbf{s}^{(j)}\right)\mathcal{H}^{(j+1)}-\mathcal{M}\right\|^2 \tag{21}$$

where $\mathcal{S}$ is the (discrete) domain of $\mathbf{s}$. The optimizations in (20) and (21) are linear least-squares problems having the solutions

$$\begin{aligned}\hat{\mathbf{m}}^{(j+1)} &= \frac{1}{T-L}\sum_{n=L}^{T-L}\left[\mathbf{y}(n)-\sum_{k=1}^{K}\sum_{l=0}^{L}\hat{\mathbf{h}}_k^{(j+1)}(n;l)s_k^{(j)}(n-l)\right]\\ \hat{\mathcal{H}}^{(j+1)} &= \mathcal{T}^{\dagger}\left(\mathbf{s}^{(j)}\right)\left[\mathbf{Y}-\mathcal{M}^{(j)}\right].\end{aligned} \tag{22}$$

whereas the optimization in (19) can be achieved by using the (vector) Viterbi algorithm. Since the above iterative procedure involving (19)-(21) decreases the cost at every iteration, one achieves a local minimum of the nonlinear least-squares cost (local maximum of DML function).

4. SIMULATION EXAMPLE

We conclude with a simulation example dealing with a two-transmitter and two-receiver scenario ($K = N = 2$). We assume both users have the same transmitted power in training and information parts. Considering a random frequency-selective Rayleigh fading channel, we took $L = 2$ in (1) with $\mathbf{h}_k(n;l)$ satisfying Jakes' model for each tap with mutually independent and identically distributed taps with unit variance. We consider a system with carrier frequency of 2GHz, data rate of 40kBd (kilo-Bauds), therefore, $T_s = 25\times 10^{-6}$sec., and Doppler spread $f_d = 50$ or 100 Hz. We emphasize that the DPS-BEM is used only for processing at the receiver; the random channels are generated by Jakes' model, not the DPS-BEM in (2). The additive noise was zero-mean complex white Gaussian ($\mathbf{m} = \mathbf{0}$). The (receiver) SNR refers to the energy per bit per user over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy. Information sequences as well as superimposed training sequences were BPSK (binary). We took $\tilde{P} = 7$ and $P = 14$ in (**H4**). The training sequence for the first user (before scaling by σ_{c1}) is

$$\{c_1(n)\}_{n=0}^{13} = \{1,-1,-1,1,1,1,-1,1,-1,-1,1,1,1,-1\},$$

(a repetition of an m-sequence of period $\tilde{P} = 7$), and $c_2(n)$ satisfies (5) with $c'_{m2} = c'_{m1}$. The average transmitted power σ_{ck}^2 in $c_k(n)$ was 0.3 of the power in $b_k(n)$, leading to a training-to-information power ratio (TIR) of 0.3.

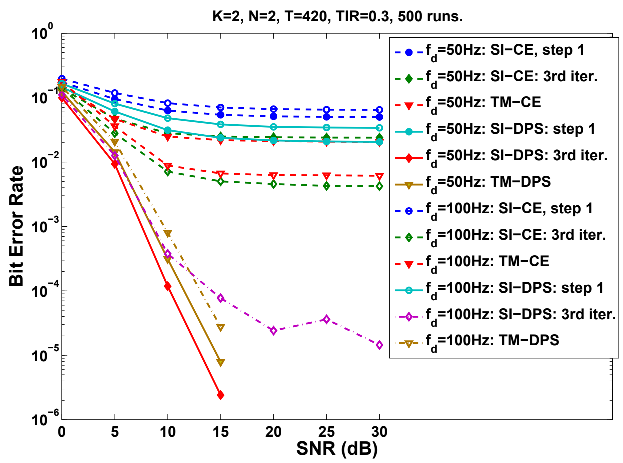

Fig. 1. *BER vs SNR. SI-CE: CE-BEM-based superimposed training; SI-DPS: DPS-BEM-based superimposed training; TM-CE: CE-BEM-based time-multiplexed training; TM-DPS: DPS-BEM-based time-multiplexed training; step 1: first-order statistics-based approach; 3rd iter.: 3rd iteration of the DML approach.*

The results for a data block length of $T = 420$ bits are shown in Figs. 1 and 2, for BER and normalized channel MSE (NCMSE) respectively. The NCMSE is defined as

$$\text{NCMSE} = \frac{(KM_r)^{-1} \sum_{i=1}^{M_r} \sum_{k=1}^{K} \sum_{n=0}^{T-1} \sum_{l=0}^{L} \left\| \hat{\mathbf{h}}_k^{(i)}(n;l) - \mathbf{h}_k^{(i)}(n;l) \right\|^2}{(KM_r)^{-1} \sum_{i=1}^{M_r} \sum_{k=1}^{K} \sum_{n=0}^{T-1} \sum_{l=0}^{L} \left\| \mathbf{h}_k^{(i)}(n;l) \right\|^2}$$

where $\mathbf{h}_k^{(i)}(n;l)$ is the true channel and $\hat{\mathbf{h}}_k^{(i)}(n;l)$ is the estimated channel at the i-th run, among the total M_r runs. The corresponding detection results are based on the Viterbi algorithm utilizing the estimated channel. The iterations follow our DML approach in Sec. 3. For $f_d = 50$ and $100\,\text{Hz}$, we choose the number of basis functions $Q = 3$ and 5 for CE-BEM, and $Q = 3$ and 4 for DPS-BEM, respectively.

For comparison, we consider a CE- or DPS-BEM-based periodically placed time-multiplexed training with zero-padding, following the design of [3]. We took a training block of length of $(K+1)L + K = 8$ bits with the first user's training sequence as $\left\{0, 0, \sqrt{b}, 0, 0, 0, 0, 0\right\}$, ($b = (L+1)(\sigma_{b1}^2 + \sigma_{c1}^2)$), and the second user's training is taken as $\left\{0, 0, 0, 0, 0, \sqrt{b}, 0, 0\right\}$. A data block of 27 bits is inserted between two such training blocks to form a frame of length 35 bits. Such a frame is repeated over a record length of 420 bits. Thus, we have a training-to-information bit power ratio of about 0.3, the same as the TIR for superimposed training.

As noted in [9], DPS-BEM efficiently reduces the spectral leakage induced by CE-BEM leading to a much smaller modeling error — BER and NCMSE curves both exhibit this advantage. It is also seen that the DML algorithm, whether CE- or DPS-BEM-based, significantly reduces the interference from the information sequence, which is induced by its first step, the first-order statistics-based approach. The BER performance after several DML iterations is competitive with the optimal time-multiplexed training without incurring the 30% training overhead penalty. For the first-order statistics-based estimator, we also plotted the theoretical channel MSE in Fig. 2. The theoretical expression and simulation-based MSE agree quite well verifying that the modeling error of DPS-BEM and the approximation error in (10) are rather small.

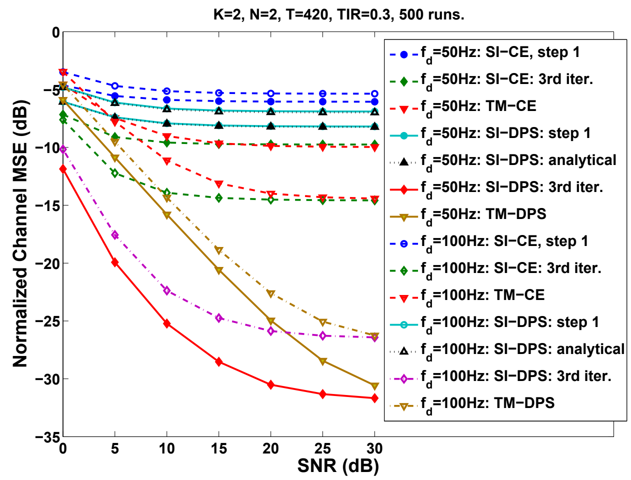

Fig. 2. *Normalized channel MSE vs SNR. SI-CE: CE-BEM-based superimposed training; SI-DPS: DPS-BEM-based superimposed training; TM-CE: CE-BEM-based time-multiplexed training; TM-DPS: DPS-BEM-based time-multiplexed training; step 1: first-order statistics-based approach; 3rd iter.: 3rd iteration of the DML approach.*

5. REFERENCES

[1] M. Ghogho, D. McLernon, E. Alamdea-Hernandez and A. Swami, "SISO and MIMO channel estimation and symbol detection using data-dependent superimposed training," in *Proc. 2005 IEEE ICASSP*, vol. III, pp. 461-464, Philadelphia, PA, March 2005.

[2] S. He and J.K. Tugnait, "Doubly-selective channel estimation using superimposed training and discrete prolate spheroidal basis models," in *Proc. 2006 IEEE GLOBECOM*, San Francisco, CA, Nov. 2006.

[3] X. Ma, G.B. Giannakis and S. Ohno, "Optimal training for block transmissions over doubly-selective fading channels" *IEEE Trans. Signal Processing*, vol. 51, pp. 1351-1366, May 2003.

[4] X. Ma and G.B. Giannakis, "Maximum-diversity transmissions over doubly-selective wireless channels" *IEEE Trans. Inform. Theory*, vol. 49, pp. 1832-1840, July 2003.

[5] X. Meng and J.K. Tugnait, "MIMO channel estimation using superimposed training," in *Proc. 2004 IEEE ICC*, pp. 2663-2667, Paris, France, June 2004.

[6] X. Meng and J.K. Tugnait, "Doubly-selective MIMO channel estimation using superimposed training," in *Proc. 2004 3rd IEEE Sensor Array & Multichannel Signal Proc. Workshop*, pp. 407-411, Barcelona, Spain, July 2004.

[7] J.G. Proakis and D.G. Manolaks, *Digital Signal Processing*, 3rd Ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.

[8] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, pp. 1371-1430, May-June 1978.

[9] T. Zemen and C.F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Processing*, vol. 53, pp. 3597-3607, Sept. 2005.