Self-Interference Suppression in Doubly-Selective Channel Estimation Using Superimposed Training

Shuangchi He and Jitendra K. Tugnait
Department of Electrical & Computer Engineering, Auburn University, Auburn, AL 36849
Email: heshuan@auburn.edu, tugnajk@eng.auburn.edu

Abstract—Channel estimation for frequency-selective time-varying channels is considered using superimposed training. We employ a discrete prolate spheroidal basis expansion model (DPS-BEM) to describe the time-varying channel. A periodic (non-random) training sequence is arithmetically added (superimposed) at low power to the information sequence at the transmitter before modulation and transmission. In existing first-order statistics-based channel estimators, the information sequence acts as interference resulting in a poor signal-to-noise ratio (SNR). In this paper a data-dependent superimposed training sequence is used to either totally or partially cancel out the effects of the unknown information sequence at the receiver on channel estimation. In total cancellation, at certain frequencies, the information-bearing components are nulled. To compensate for this information loss, we propose a partially-data-dependent (PDD) superimposed training scheme where a trade-off is made between interference cancellation and frequency integrity. An iterative method is also used to enhance channel estimation and data detection and illustrated via a simulation example.

I. INTRODUCTION

Consider a doubly-selective (time- and frequency-selective) SIMO (single-input multi-output) FIR (finite impulse response) linear channel with $N$ outputs. Let $\{s(n)\}$ denote a scalar sequence which is input to the SIMO time-varying channel with discrete-time response $\{h(n;l)\}$ ($N$-vector channel response at time $n$ to a unit input at time $n-l$). Then the symbol-rate channel output vector is given by

$$\mathbf{x}(n) = \sum_{l=0}^{L} \mathbf{h}(n;l) s(n-l).$$

(1)

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

$$\mathbf{h}(n;l) = \sum_{q=1}^{Q} \mathbf{h}_q(l) u_q(n)$$

(2)

where $u_q(\cdot)$ is the $q$-th basis function ($q = 1, \cdots ,Q$), and $N$-column vectors $\mathbf{h}_q(l)'s$ are fixed over the data block. In the complex exponential basis expansion model (CE-BEM) [1], for a record length of $T$ symbols with symbol interval $T_s$ sec., one chooses $u_q(n) = e^{j\omega_q n}$, $\omega_q := 2\pi \left(q - \frac{Q+1}{2}\right)/T$, $L := [\tau_d/T_s]$, and $Q := 2[f_dT_T] + 1$ when the underlying continuous-time channel has a delay spread of $\tau_d$ sec. and Doppler spread of $f_d$ Hz. In discrete prolate spheroidal BEM (DPS-BEM), the $i$-th DPS vector $u_i := [u_i(0), \cdots , u_i(T-1)]^T$ (called Slepian sequence in [2], which is a time-windowed (infinite) DPS sequence) is the $i$-th eigenvector of a matrix $\mathbf{C}$ [3]: $\mathbf{C} u_i = \lambda_i u_i$, where $[\mathbf{C}]_{n,m} = \sin[2\pi(n-m)f_dT_s]/\pi(n-m)$ 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 Slepian sequences $u_i(n)$ are orthonormal over the finite time interval $[0,T-1]$.

The rectangular window of the truncated DFT-based CE-BEM model introduces spectral leakage [4], [2]. 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 [2]. DPS sequences are a good alternative as a basis set to approximate bandlimited channels alleviating the spectral leakage of CE-BEM [2]. In this case one takes $Q = \lceil 2f_s T_s \rceil + 1$ [2]. The (infinite) DPS sequences have their maximum energy concentration in an interval with length $T$ while being bandlimited to $[-f_s T_s, f_s T_s]$, where $u_0(n)$ is the unique sequence that is bandlimited and most time-concentrated, $u_1(n)$ is the next sequence having maximum energy concentration among the DPS sequences orthogonal to $u_0(n)$, and so on [3]. Time-windowed DPS sequences (the Slepian sequences) are used in (2) to represent the time-variant channel over data-block

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

(3)

Our objective is to recover $s(n)$ given noisy $\{\mathbf{y}(n)\}$. In several approaches this requires knowledge of the channel impulse response. Recently a superimposed training based approach has been explored to this end where one takes

$$s(n) = b(n) + c(n),$$

(4)

where $\{c(n)\}$ is a training (pilot) sequence added (superimposed) at low power to the information sequence $\{b(n)\}$ at the transmitter before modulation and transmission. There is no loss in data transmission rate unlike the conventional time-multiplexed training, but some useful power is wasted in superimposed training. Periodic superimposed training has been discussed in [5], [6] for time-invariant channels, in [7] and [8] for time-varying channels based on CE-BEM, and in [9] for time-varying channels based on DPS-BEM. Ref.
[2] is the first to use DPS-BEM for doubly-selective channel estimation using time-multiplexed training.

In superimposed training-based channel estimation, the contribution of the unknown information sequence acts as interference (called self-interference in this paper) resulting in a lower SNR for channel estimation [10]. In [6], using a block transmission approach, a data-dependent superimposed training sequence is used at the transmitter to cancel out the interference from the information sequence at the receiver for time-invariant channel estimation. This result is extended to CE-BEM-based doubly-selective channels in [8] using a block transmission approach.

Objectives and Contributions: Based on the channel estimator proposed in [8] using data-dependent superimposed training, we extend the CE-BEM-based approach of [8] to that based on DPS-BEM. We estimate the doubly-selective channel \( h(n; l) \) using the knowledge of superimposed training, as well as the first-order statistics of the noisy observations. However, the interference part canceled by data-dependent training at certain frequencies is information-bearing, removal of which at the transmitter results in information loss before transmission, and it causes a loss in information detection performance at the receiver. To remedy this we propose a partially-data-dependent (PDD) scheme to strike a compromise between interference elimination and information integrity. Instead of total annihilation, the interference is suppressed only partially, so that a “small” portion of this information-bearing component remains for “future” recovery of the information sequence at the receiver. An iterative deterministic maximum likelihood (DML) approach is also considered to enhance the first-order statistics-based estimator, where in the subsequent iterations, the estimate of the information sequence is exploited to enhance channel estimation and recover the suppressed information components.

Notation: Superscripts \( H, \ast, T, \) and \( \dagger \) denote the complex conjugate transpose, complex conjugation, transpose, and Moore-Penrose pseudo-inverses, respectively. \( \delta (\cdot) \) is the Kronecker delta function and \( I_N \) is the \( N \times N \) identity matrix. The symbol \( \otimes \) denotes the Kronecker product, and \( \text{tr} (A) \) is the trace of a square matrix \( A \). The \( (n,m) \)-th entry of a matrix \( C \) is denoted by \( C_{n,m} \).

II. FIRST-ORDER STATISTICS-BASED ESTIMATOR USING DPS-BEM [9]

Here we summarize the first-order statistics-based estimator proposed in [9] using DPS-BEM. Assume the following:

1. The time-varying channel satisfies (2). The information sequence \( \{b(n)\} \) is zero-mean, white, with variance \( E\{b(n)^2\} = \sigma_b^2 \). The measurement noise \( \{v(n)\} \) is zero-mean, white, uncorrelated with \( \{b(n)\} \), with autocorrelation \( E\{v(n+\tau)v^H(n)\} = \sigma_v^2 \delta(\tau) \). The superimposed training sequence \( c(n) = c(n+mP) \) exists for any \( m \) and \( P \) is a non-random periodic sequence with period \( P \).

Since \( c(n) \) is \( P \)-periodic, we have

\[
c(n) = \sum_{m=0}^{P-1} c_m e^{j\alpha_m n}, \quad \forall n
\]

where \( c_m := \sum_{n=0}^{P-1} c(n) e^{-j\alpha_m n}/P \) and \( \alpha_m := 2\pi m/P \). By (1)–(5),

\[
E\{v(n)\} = \sum_{q=1}^Q \sum_{m=0}^{P-1} \left[ \sum_{l=0}^L c_{m,q} h_q(l) e^{-j\alpha_m l} \right] u_q(n) e^{j\alpha_m n}.
\]

Define

\[
A := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-j\alpha_1} & e^{-j\alpha_2} & \cdots & e^{-j\alpha_L} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j\alpha_{P-1}} & e^{-j\alpha_{P-2}} & \cdots & e^{-j\alpha_{P-L}} \end{bmatrix}
\]

\[
D_m := [d_{m,1}^T, \ldots, d_{m,Q}^T]^T, \quad \mathcal{D} := \left[ D_0^H, \ldots, D_{P-1}^H \right]^H,
\]

\[
H_l := [h_l^T(l), \ldots, h_l^T(L)]^T, \quad \mathcal{H} := [H_0^H, \ldots, H_L^H]^H,
\]

\[
C := \text{diag} \{ c_0, \ldots, c_{P-1} \} \otimes I_{NQ}.
\]

By the definition of \( d_{mq} \) in (6), it then follows that

\[
C \mathcal{H} = \mathcal{D}.
\]

It is shown in [7] that if \( P \geq L+1, \text{rank}(C) = NQ(L+1) \). Hence, we can determine the \( h_q(l) \)'s uniquely. In [9], an estimate \( \hat{d}_{mq} \) of \( d_{mq} \) is given as

\[
\hat{d}_{mq} = \sum_{n=0}^{T-1} V(n) u_q(n) e^{-j\alpha_m n}
\]

using the fact that the (infinite) DPS sequences are band-limited to the normalized frequency range \([-f_d T_s, f_d T_s]\), so that the time-limited DPS sequences obtained by rectangular windowing over \([0, T-1]\), approximately satisfy

\[
\sum_{n=0}^{T-1} u_{n'}(n) u_q(n) e^{j(\alpha_{n'} - \alpha_m)n} \approx \delta(m' - m) \delta(q' - q)
\]

if \( f_d T_s \ll 1/P \) and \( T \) is a multiple of \( P \) or \( T \) is “large”.

Define \( \hat{\mathcal{H}} \) as in (8) with \( d_{mq} \) replaced with \( \hat{d}_{mq} \). The estimation of the channel coefficients is then given by

\[
\hat{\mathcal{H}} = (C^H C)^{-1} C^H \hat{\mathcal{D}}.
\]

The channel estimate is then given by

\[
\hat{h}(n; l) = \sum_{q=1}^Q \hat{h}_q(l) u_q(n).
\]

Remark 1. As noted in [9], if the mean of the noise \( v(n) \) is unknown, say

\[
E\{v(n)\} = \mathbf{m},
\]
one should omit the first row (corresponding to $\alpha_0$) of $V$ in (7) (denote the resulting matrix by $V'$), and the block $D_0$ from $D$ in (8) (denote the resulting matrix by $\hat{D}$). Define

$$\hat{C} := \left(\text{diag} \left\{ c_1, \cdots, c_{p-1} \right\} \right) V'. $$

Then we have $\hat{C} \bar{H} = \hat{D}$ and

$$\bar{H} = (\hat{C}^H \hat{C})^{-1} \hat{C}^H \hat{D},$$

(15)

where $\hat{D}$ is also acquired by (10). For identifiability, we now need $P \geq L + 2$. All our subsequent results hold true if appropriate substitutions are used.

III. DATA-DEPENDENT SUPERIMPOSED TRAINING-BASED SOLUTION

Consider (10). It has contributions from the information sequence $\{b(n)\}$ unknown at the receiver, the superimposed training $\{c(n)\}$ known at the receiver, and the noise $v(n)$. It follows from (H1), (1)–(4) and (10), that

$$\hat{d}_{mq} = \sum_{n=0}^{T-1} \left[ E \{ y(n) \} + \sum_{l=0}^{L} h(n;l) b(n-l) + v(n) \right] \times u_q(n) e^{-j\alpha_m n}.$$

(16)

For a given training period $P$, we pick the block size $T$ such that $T = KP$ where $K \geq Q$ is an integer. Then by (H1), (6), (11), and (16)

$$E\left[ \hat{d}_{mq} \right] = d_{mq} = \sum_{n=0}^{T-1} E \{ y(n) \} u_q(n) e^{-j\alpha_m n}.$$  

(17)

Define $w_{mq} := \sum_{n=0}^{T-1} v(n) u_q(n) e^{-j\alpha_m n}$, which is zero-mean and by (H1) and (11)

$$E \left\{ w_{mq'} w_{mq}^H \right\} = \sigma_v^2 I_N \delta(m'-m) \delta(q'-q).$$

(18)

Thus

$$\hat{d}_{mq} = d_{mq} + s_{mq} + w_{mq}$$

(19)

where

$$s_{mq} := \sum_{n=0}^{T-1} \sum_{l=0}^{L} h(n;l) b(n-l) u_q(n) e^{-j\alpha_m n}.$$  

(20)

Clearly, the information sequence’s contribution, given by $s_{mq}$, interferes with the estimation of $d_{mq}$, hence with the channel estimation from the observations (see (12)).

Consider the DFT of the information sequence $\{b(n)\}_{n=0}^{T-1}$:

$$b_r := \frac{1}{T} \sum_{n=0}^{T-1} b(n) e^{-j\omega_r n} \text{ and } b(n) = \sum_{r=0}^{T-1} b_r e^{j\omega_r n},$$

(21)

where $\omega_r := \frac{2\pi r}{T}$. Then the information-induced interference $s_{mq}$ in (20) can be expressed as

$$s_{mq} = \sum_{q'=1}^{Q} \sum_{l=0}^{L} \sum_{r=0}^{T-1} \left[ h_{q'}(l) e^{-j\omega_r} b_r \right] \sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j2\pi(r-mK)n/T}. $$

Again we exploit the approximate bandlimitedness of the time-limited DPS sequences to obtain

$$\sum_{n=0}^{T-1} u_{q'}(n) u_q(n) e^{j2\pi(r-mK)n/T} \approx 0$$

(22)

for $|r-mK| \geq Q + k$, where $k$ is an integer and $(Q + k)/T > 2f_d T_s$. Therefore, if we set $b_r = 0$ for those $r$’s belonging to a set

$$\Omega := \{ r | (Q + k) + mK \leq r \leq (Q + k) + mK, m = 0, \cdots, P-1 \},$$

(23)

then $s_{mq} = 0$. We do so by modifying $\{c(n)\}$ based on $\{b(n)\}$ (at the transmitter). Define a “self-interference” sequence

$$b_c(n) := \sum_{r=0}^{T-1} b_r e^{j\omega_r n}$$

(24)

and a data-dependent superimposed training sequence $\{\tilde{c}(n)\}_{n=0}^{T-1}$ such that

$$\tilde{c}(n) := c(n) - b_c(n).$$

(25)

Note that $\{\tilde{c}(n)\}$ is no longer periodic with period $P$. At the transmitter, we now send

$$s(n) = \tilde{c}(n) + b(n) = c(n) + [b(n) - b_c(n)].$$

The model (1)–(4) holds with $c(n)$ replaced with $\tilde{c}(n)$. By construction, the DFT of $\{b(n) - b_c(n)\}_{n=0}^{T-1}$ vanishes at frequency $\omega_r$ for $r$ in the set $\Omega$. Also the DFT of $\{b(n) - b_c(n)\}_{n=0}^{T-1}$ vanishes at frequencies in the set $\Omega$ provided that a cyclic prefix of length $M \geq L$ is used. A cyclic prefix of length $M$ is added at the transmitter by choosing

$$s(-i) = s(T - i), \quad i = 1, \cdots, M \geq L$$

where $s(i) = \tilde{c}(i) + b(i)$. This allows the linear convolution in (1) to be equal to the circular convolution (implicit in the DFT operation) over the block length $n = 0, \cdots, T-1$.

We summarize our data-dependent channel estimation solution as follows:

1) At the transmitter, we are given information sequence over a block as $\{b(n)\}$ for $0 \leq n \leq T - 1$ with $T$ chosen as $T = KP, K \geq Q$. Calculate the DFT by (21).

2) To eliminate interference with channel estimation at the receiver, we need to set $b_r$’s to zero for $r \in \Omega$. Define $\{b_r(n)\}$ as in (24).

3) Define the data-dependent superimposed training $\tilde{c}(n)$ as in (25). Use a cyclic prefix of length $M \geq L$ and transmit.

4) The channel estimation given in (12) stays the same for data-dependent superimposed training, because we still use periodic $\{c(n)\}$ at the receiver for symbol detection, and we do not know $b_c(n)$ or $b(n)$ at the receiver. Now there is no contribution of $\{b(n)\}$ to $\hat{d}_{mq}$ for $0 \leq m \leq P - 1$ and $1 \leq q \leq Q$. 

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.
A. Data Detection

Now the “information sequence” is \( \{ b(n) - b_c(n) \} \) whereas we are interested in \( \{ b(n) \} \). We will follow an iterative solution, similar to the time-invariant results of [6] and time-varying results of [9]. The first step in our solution is to use the estimated channel to detect \( \{ b(n) \} \) via Viterbi algorithm (ignoring \( b_c(n) \) but accounting for the known \( \{ c(n) \} \)). Use the detected \( \{ b(n) \} \) to estimate \( \{ b_c(n) \} \), and iterate the detection procedure (but not channel estimation) with known \( \{ c(n) \} \) and estimate \( \{ b_c(n) \} \) from the previous iteration. Note that the receiver only knows \( \{ c(n) \} \) and has knowledge of how \( \{ b_c(n) \} \) is generated from \( \{ b(n) \} \), but does not know \( \{ b(n) \} \).

IV. PARTIALLY DATA-DEPENDENT (PDD) SUPERIMPOSED TRAINING

If the channel has bandwidth \( B_c \), since the training sequence is \( P \)-periodic, at the receiver the training components of the received signal only occur at \( P \) distinct frequency intervals of bandwidth \( B_c \) each (suppose \( 1/P \gg B_c \) so that there is no overlap among these intervals). In the data-dependent superimposed training method, by setting \( b_r = 0 \) for \( r \in \Omega \) before transmission, we discard the frequency components of the information sequence at those intervals and their neighborhoods to avoid aliasing, so that information-induced self-interference is eliminated. By (23), however, the information contained at those \( P^2(2Q+2k+1) \) (among total \( T \)) frequencies is also discarded. Though it may be partially recovered by other properties (e.g., the finite alphabet of the channel), it can cause severe detection errors under severe frequency loss.

For the first-order statistics-based estimator we discussed in Sec. II, the sequence \( \{ b_c(n) \} \) acts as interference in channel estimation, however, it also bears information so that it should not be simply discarded. Given \( \{ b_c(n) \} \) as in (24), we now transmit

\[
s(n) = c(n) + b(n) - (1 - \gamma) b_c(n) \tag{26}
\]

at the transmitter, where \( 0 \leq \gamma \leq 1 \) is the interference factor. When \( \gamma = 1 \), the information sequence \( \{ b(n) \} \) stays intact, corresponding to the first-order statistics-based estimator in Sec. II. When \( \gamma = 0 \), the interference-induced frequency components \( b_r \)'s \( r \in \Omega \) are totally annihilated, corresponding to the data-dependent solution described in Sec. III. If \( 0 < \gamma \ll 1 \), which is the partially-data-dependent (PDD) case, at each \( r \in \Omega \) the frequency component \( b_r \) is reduced to \( \gamma b_r \). Then the self-interference will be “sufficiently” suppressed when conducting channel estimation, while the frequency components at \( r \in \Omega \) remain “partially” intact. Note that in this PDD method, the self-interference is not completely removed, so that the channel estimation is not as accurate as that in the “completely” data-dependent solution. But since no information-bearing frequencies are nulled out, the information contained there can be recovered in data reception.

Our PDD superimposed training-based channel estimation follows the data-dependent solution described in Sec. III, except that we define the PDD superimposed training sequence as

\[
\tilde{c}(n) := c(n) - (1 - \gamma) b_c(n)
\]

instead of (25), for a preselected interference factor \( \gamma \).

V. PERFORMANCE ANALYSIS

We wish to evaluate the mean-square error (MSE) in channel estimation using PDD superimposed training when the true channel follows (2):

\[
\text{MSE} = \frac{1}{T} \sum_{n=0}^{T-1} \sum_{l=0}^{L} E \left\{ \left\| h(n;l) - \hat{h}(n;l) \right\|^2 \right\}, \tag{27}
\]

where \( h(n;l) \) follows (2) and \( \hat{h}(n;l) \) follows (13). We make the following assumption about the channel \( h(n;l) \):

- \( \text{(H2)} \) The time-varying channel \( \{ h(n;l) \} \) is zero-mean, complex Gaussian with \( E \{ h(n;l) h^H(n;l) \} = \sigma^2_h \mathbf{I}_N \) and \( E \{ h(n_1;l_1) h^H(n_2;l_2) \} = 0 \), for \( l_1 \neq l_2, n_1, n_2 \).

Note that in PDD superimposed training, (19) still holds if \( s_{mq} \) is revised as

\[
s_{mq} := \sum_{n=0}^{T-1} \left\{ \sum_{l=0}^{L} h(n;l) [b(n - l) - (1 - \gamma) b_c(n)] \right\} \times u_q(n) e^{-js_0 n n}.
\]

By \( \text{(H1)}, \text{(H2)}, \) (11), and (24), we have

\[
E \{ s_{mq} s_{m'q'}^H \} = \gamma^2 (L + 1) \sigma^2_h \sigma^2_q \mathbf{I}_N \delta(m' - m) \delta(q' - q).
\]

Since \( \mathbf{C} \) is full column-rank when \( P \geq L + 1 \), by (9) and (17), we have \( E \{ \mathbf{D} \} = \mathbf{D} \) and \( E \{ \mathbf{H} \} = \mathbf{H} \). Then from (12)

\[
cov \{ \mathbf{H}, \mathbf{H} \} := E \{ [\mathbf{H} - \mathbf{H}] [\mathbf{H} - \mathbf{H}]^H \}
\]

\[
= (\mathbf{C} \mathbf{H})^{-1} \mathbf{C} \mathbf{H} \text{cov} \{ \mathbf{D}, \mathbf{D} \} \mathbf{C} (\mathbf{C} \mathbf{H})^{-1}.
\]

Since

\[
E \{ [\mathbf{d}_{mq} - d_{mq}] [\mathbf{d}_{m'q'} - d_{m'q'}]^H \}
\]

\[
= E \{ s_{mq} s_{m'q'}^H \} + E \{ w_{m'q} w_{m'q'}^H \},
\]

by (18) and (28) we have

\[
cov \{ \mathbf{D}, \mathbf{D} \} = \left[ \gamma^2 (L + 1) \sigma^2_h \sigma^2_q + \sigma^2_v \right] \mathbf{I}_{NPQ}.
\]

Substitute (30) for (29)

\[
cov \{ \mathbf{H}, \mathbf{H} \} = \left[ \gamma^2 (L + 1) \sigma^2_h \sigma^2_q + \sigma^2_v \right] (\mathbf{C} \mathbf{H})^{-1}.
\]

Using the orthonormality of the Slepian sequences, the MSE in channel estimation (27) is given by

\[
\text{MSE} = \frac{1}{T} E \left\{ \sum_{l=0}^{L} \sum_{q=1}^{Q} \left\| h_q(l) - \hat{h}_q(l) \right\|^2 \right\}
\]

\[
= \frac{\gamma^2 (L + 1) \sigma^2_h \sigma^2_q + \sigma^2_v}{T} \text{tr} \{ (\mathbf{C} \mathbf{H})^{-1} \}.
\]

The MSE of channel estimation of the first-order statistics-based estimator in Sec. II is given by (31) for \( \gamma = 1 \), and that
of the data-dependent solution in Sec. III is corresponding to \( \gamma = 0 \). For a PDD scheme with \( 0 < \gamma \ll 1 \), the interference is significantly suppressed.

VI. DETERMINISTIC MAXIMUM LIKELIHOOD (DML) APPROACH [9]

For data detection at the receiver, now the “information sequence” is \( \{ b(n) \} = (1 - \gamma) b_r(n) \) \((0 \leq \gamma \leq 1)\), while we are interested in \( \{ b(n) \} \). We can first use the estimated channel to detect \( \{ b(n) \} \) via Viterbi algorithm (ignoring \((1 - \gamma) b_r(n)\) but accounting for the known \( \{ c(n) \} \)). Since the training and information sequences pass through an identical channel, this fact can be exploited to recover the suppressed frequency components \( b_r(n) \) \((r \in \Omega)\), as well as enhance channel estimation, in an iterative way. The details are similar to those in [9, Sec. V], hence are omitted.

VII. SIMULATION EXAMPLES

We consider a random frequency-selective Rayleigh fading channel. We took \( N = 1 \) and \( L = 2 \) in (1) with \( h(n; \ell) \) as in (H2) satisfying Jakes’ model. We consider a system with carrier frequency of 2 GHz, data rate of 40kBd (kilo-Bauds), therefore, \( T_s = 25 \mu s \), and a varying Doppler spread \( f_d \) in the range of 50 Hz to 200 Hz (corresponding to a maximum mobile velocity from 27 to 108 km/h). 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).

Additive noise was zero-mean complex white Gaussian \((\mathbf{m} = 0)\). The (receiver) SNR refers to the energy per bit over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy. The information sequence was BPSK (binary). We took the superimposed training sequence of period \( P = 4 \) in (H1) as \( c(n) = \sigma_p e^{j\pi n(n+\nu)/P} \) where \( \nu = 1 \) if \( P \) is odd and \( \nu = 2 \) if \( P \) is even, as in [5].

The results for a record length of \( T = 400 \) symbols are shown in Fig. 1 through Fig. 6 for various Doppler spreads and SNR’s. All results are based on 500 Monte-Carlo runs. The normalized channel MSE (NCMSE) is defined as

\[
\text{NCMSE} = \frac{\sum_{i=1}^{M} \sum_{n=0}^{T-1} \sum_{\ell=0}^{L} \left\| \hat{h}^{(i)}(n; \ell) - h^{(i)}(n; \ell) \right\|^2}{\sum_{i=1}^{M} \sum_{n=0}^{T-1} \sum_{\ell=0}^{L} \left\| h^{(i)}(n; \ell) \right\|^2}
\]

where \( h^{(i)}(n; \ell) \) is the true channel and \( \hat{h}^{(i)}(n; \ell) \) is the...
estimated channel at the \(i\)-th run, among total \(M_r\) runs. For \(f_d = 50, 100, 200\) Hz, we chose the number of basis functions \(Q = 3, 4, 6\), respectively. We took \(k = -1\) in (23), for which even for the “completely” data-dependent training scheme, the interference cannot be totally removed, but the information loss is comparatively mild.

For the superimposed training-based estimators, the average transmitted power \(\sigma_c^2\) in \(c(n)\) was 0.15 of the power in \(b(n)\), leading to a training-to-information power ratio (TIR) of 0.15. The “non-data-dependent” transmission scheme described in Sec. II (denoted by NDD in the figures), the “completely” data-dependent training scheme in Sec. III (denoted by DD in the figures), and the PDD scheme with the interference factor \(\gamma = 0.2\) are considered. We also investigate the non-data-dependent training scheme for \(TIR = 0.3\) — assigning more power to training in this scheme improves data detection, provided only a small portion of power is allocated to training. (However, for the “completely” data-dependent scheme, \(TIR = 0.15\) is around the optimal value for minimum detection errors, since the interference is greatly reduced, and more power should be allocated to the information sequence to increase the effective SNR.) At the receiver, DML iterations follow the data detection scheme we described at the beginning of Sec. VI (denoted by “step 1” in the figures). We show bit error rates (BER) and NC MSE for this first step, and the results after three iterations. For comparison, we consider a DPS-BEM-based periodically placed time-multiplexed training with zero-padding (denoted by TM in the figures), patterned after the scheme proposed in [11]. We took a training block of length \(2L + 1 = 5\) symbols with training sequence \(\{0, 0, \sqrt{(L+1)(\sigma_b^2 + \sigma_c^2)}, 0, 0\}\) and a data block of length 35 leading to a frame of length 40 symbols. This frame was repeated over a record length of 400 symbols. Thus, we have a training-to-information bit ratio of approximately 0.15.

The DML algorithm significantly improves the performance of each estimator. It is seen that the data-dependent scheme yields the best channel estimators. The PDD scheme has a distinct advantage over it in data detection, for the information loss has been effectively reduced. There is no information loss in the non-data-dependent scheme, but it needs more training power to achieve satisfactory channel estimation behavior, which decreases the effective SNR for data detection at the receiver. Whether before or after iterations, the PDD scheme performs the best in data detection among the superimposed training-based schemes. The BER performance of the PDD scheme after several DML iterations is competitive with the “optimal” time-multiplexed training, without incurring any training overhead penalty.

**REFERENCES**


