Doubly Selective Channel Estimation Using Superimposed Training and Exponential Bases Models

Jitendra K. Tugnait,[1] Xiaohong Meng,[1, 2] and Shuangchi He[1]

[1] Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849, USA
[2] Department of Design Verification, MIPS Technologies Inc., Mountain View, CA 94043, USA

Channel estimation for single-input multiple-output (SIMO) frequency-selective time-varying channels is considered using superimposed training. The time-varying channel is assumed to be described by a complex exponential basis expansion model (CE-BEM). A periodic (nonrandom) training sequence is arithmetically added (superimposed) at a low power to the 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 CE-BEM and only the first-order statistics of the data. 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 SIMO channel and the information sequences sequentially, based on CE-BEM. Three illustrative computer simulation examples are presented including two where a frequency-selective channel is randomly generated with different Doppler spreads via Jakes' model.

1. INTRODUCTION

Consider a time-varying SIMO (single-input multiple-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 impulse response $\{\mathbf{h}(n; l)\}$ (N-vector channel response at time n to a unit input at time $n - l$). The vector channel may be the result of multiple receive antennas and/or oversampling at the receiver. Then the symbol-rate, channel output vector is given by

$$\mathbf{x}(n) := \sum_{l=0}^{L} \mathbf{h}(n; l)s(n - l). \tag{1}$$

In a complex exponential basis expansion representation [4] it is assumed that

$$\mathbf{h}(n; l) = \sum_{q=1}^{Q} \mathbf{h}_q(l)e^{j\omega_q n}, \tag{2}$$

where N-column vectors $\mathbf{h}_q(l)$ (for $q = 1, 2, \ldots, Q$) are time-invariant. Equation (2) is a basis expansion of $\mathbf{h}(n; l)$ in the time variable n onto complex exponentials with frequencies $\{\omega_q\}$. The noisy measurements of $\mathbf{x}(n)$ are given by

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

Equation (2) is the complex-exponential basis expansion model (CE-BEM).

A main objective in communications is to recover $s(n)$ given noisy $\{\mathbf{y}(n)\}$. In several approaches this requires knowledge of the channel impulse response [11, 19]. In conventional training-based approaches, for time-varying channels, one has to send a training signal frequently and periodically to keep up with the changing channel [7]. This wastes resources. An alternative is to estimate the channel based solely on noisy $\mathbf{y}(n)$ exploiting statistical and other properties of $\{s(n)\}$ [11, 19]. This is the blind channel estimation approach. More recently a superimposed training-based approach has been explored where one takes

$$s(n) = b(n) + c(n), \tag{4}$$

where $\{b(n)\}$ is the information sequence and $\{c(n)\}$ is a training (pilot) sequence added (superimposed) at a low power to the information sequence at the transmitter before modulation and transmission. There is no loss in information rate. On the other hand, some useful power is wasted in superimposed training which could have otherwise been allocated to the information sequence. Periodic superimposed training for channel estimation via first-order statistics for SISO systems have been discussed in [9, 16, 21] (and references therein) for time-invariant channels, and in [17] (a conference version of Section 2 of this paper) for both time-invariant and time-varying (CE-BEM based) channels.

CE-BEM representation/approximation of doubly selective channels have been used in [1, 2, 4–7, 15], among others. Reference [7] deals with time-multiplexed training sequence design for block transmissions. In this paper we only deal with serial transmissions. In [5], a semiblind approach is considered with time-multiplexed training with serial transmissions and at least two receive antennas. In this paper our results hold even with one receive antenna. Reference [2] deals with time-varying equalizer design given CE-BEM representation.

Reference [3] appears to be the first to use (periodic) superimposed training for SISO time-invariant channel estimation. Periodic training allows for use of the first-order statistics (time-varying mean) of the received signal. Since blind approaches cannot resolve a complex scaling factor ambiguity, they require differential encoding/decoding resulting in an approximately 3 dB SNR loss. It was noted in [3] that power loss in superimposed training would be typically much less than 3 dB. Furthermore, it was also noted in [3] that identifiability conditions for superimposed training-based methods are much less stringent than that for blind approaches. As noted earlier periodic superimposed training for channel estimation via first-order statistics for SISO systems has been discussed in [17] for both time-invariant and time-varying (CE-BEM based) channels. While in principle aperiodic superimposed training can also be used, periodic training allows for a much simpler algorithm; for instance, for CE-BEM channels, relation (13) leads to (19) (see Section 2) which allows for a "decoupled" estimation of the coefficients $\mathbf{d}_{mq}$ (see (10)) from data. In the CE-BEM model the exponential basis functions are orthogonal over the record length. When we use periodic training with appropriately selected period in relation to the record length, the "composite" basis functions ($e^{j\omega_{mq}n}$ in Section 2) are still orthogonal, leading to (13). However, there does not exist any relative advantage or disadvantage between periodic and aperiodic superimposed training when using the iterative approach to joint channel and information sequence estimation discussed in Section 3. In the simulations presented in this paper we used an m-sequence (maximal length pseudorandom binary sequence) as superimposed training sequence. While there exist a large class of periodic training sequences which are periodically white and/or optimal in some sense (see [9]), some of them do not have a peak-to-average power ratio of one and some of them do not have finite alphabet, whereas an m-sequence has finite (binary) alphabet and unity peak-to-average power ratio.

As noted earlier, compared to periodically inserted time-multiplexed training (as in [7]), there is no loss in data transmission rate in superimposed training. However, there may be an increase in bit-error rate (BER) because of an SNR loss due to power allocated to superimposed training. Our simulation comparisons show that at "low" SNRs we also have a BER advantage (see Example 3 in Section 4). In semi-blind approaches (such as that in [5]), there is periodically inserted time-multiplexed training but one uses the nontraining-based data also to improve the training-based results: it uses a combination of training and blind cost functions. While [5]

needs at least two receive antennas, in this paper our results hold even with one receive antenna; besides, in [5] there is still a loss in data transmission rate owing to the presence of time-multiplexed training.

In [17] a first-order statistics-based approach for time-invariant channel estimation using periodic superimposed training has been presented. This approach is further analyzed and enhanced in [18] where a performance analysis has been carried out, and issues such as frame synchronization and training power allocation have been discussed. Both these papers do not deal with time-varying channels; moreover, they do not discuss any iterative approach to joint channel and information sequence estimation even in the context of time-invariant channels.

Objectives and contributions

In this paper, we first present and extend the first-order statistics-based approach of [17] for time-varying (CE-BEM based) channels. Then we extend the first-order statistics-based solution to an iterative approach to joint channel and information sequence estimation, based on CE-BEM, using Viterbi detectors. The first-order statistics-based approach views the information sequence as interference whereas in the iterative joint estimation version it is exploited to enhance channel estimation and information sequence detection. All results in this paper are developed for an SIMO formulation since everything developed for an SISO system carries over to an SIMO model in a straightforward fashion. However, all our simulations are presented for an SISO system (for simplicity of presentation).

Notation

Superscripts H, T, and $\dagger$ denote the complex conjugate transpose, the transpose and the Moore-Penrose pseudoinverse operations, respectively. $\delta(\tau)$ is the Kronecker delta and I_N is the $N \times N$ identity matrix. The symbol $\otimes$ denotes the Kronecker product. The superscript $*$ denotes the complex conjugation operation.

1.1. CE-BEM representation

We now briefly discuss the CE-BEM representation of time-varying communications channels, following [4] and particularly [6], to consider practical situations where the basis frequencies ω_q's would be known a priori. Consider a time-varying (e.g., mobile wireless) channel with complex baseband, continuous-time, received signal $x(t)$ and transmitted complex baseband, continuous-time information signal $s(t)$ (with symbol interval T_s seconds) related by $h(t; \tau)$ which is the time-varying impulse response of the channel (response at time t to a unit impulse at time $t - \tau$). Let τ_d denote the (multipath) delay-spread of the channel and let f_d denote the Doppler spread of the channel. If $x(t)$ is sampled once every T_s seconds (symbol rate), then by [6], for $t = nT_s + t_0 \in [t_0, t_0 + TT_s)$, the sampled signal $x(n) := x(t)|_{t=nT_s+t_0}$ has the

representation

$$x(n) = \sum_{l=0}^{L} h(n; l) s(n - l), \qquad (5)$$

where

$$h(n; l) = \sum_{q=1}^{Q} h_q(l) e^{j\omega_q n}, \qquad L := \left\lfloor \frac{\tau_d}{T_s} \right\rfloor, \qquad (6)$$

$$\omega_q = \frac{2\pi}{T} \left(q - \frac{1}{2} - \frac{Q}{2} \right), \quad Q := 2\lceil f_d T T_s \rceil + 1. \qquad (7)$$

This is a scenario where the CE-BEM representation is appropriate. The above representation is valid over a duration of TT_s seconds (T samples). Equation (1) arises if we follow (5) and consider an SIMO model arising due to multiple antennas at the receiver. Although discussed in the context of OFDM, in [12] it is shown that finite-duration observation window effects compromise the accuracy of CE-BEM, that is, CE-BEM is "accurate" only as $T \to \infty$. One could try to improve the CE-BEM efficacy by explicitly incorporating time-domain windowing effects (as in [12]). Such modifications are outside the scope of this paper. We do note that in [8], alternative models (such as polynomial bases models) coupled with CE-BEM have been used to improve the modeling results.

2. A FIRST-ORDER STATISTICS-BASED SOLUTION

It is based on CE-BEM. Assume the following:

(H1) the time-varying channel $\{\mathbf{h}(n; l)\}$ satisfies (2) where the frequencies ω_q ($q = 1, 2, \ldots, Q$) are distinct and known with $\omega_q \in [0, 2\pi)$. Also $N \geq 1$. For some $\overline{q}$ ($1 \leq \overline{q} \leq Q$), we have $\omega_{\overline{q}} = 0$;

(H2) the information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = 1$;

(H3) the measurement noise $\{\mathbf{v}(n)\}$ is *nonzero-mean* ($E\{\mathbf{v}(n)\} = \mathbf{m}$), white, uncorrelated with $\{b(n)\}$, with $E\{[\mathbf{v}(n + \tau) - \mathbf{m}][\mathbf{v}(n) - \mathbf{m}]^H\} = \sigma_v^2 I_N \delta(\tau)$. The mean vector $\mathbf{m}$ may be unknown;

(H4) the superimposed training sequence $c(n) = c(n + mP)$ for all m, n is a nonrandom periodic sequence with period P.

For model (7), we have $\overline{q} = (Q+1)/2$. Negative values of ω_q's in (7) are to be interpreted as positive values after a modulo 2π operation, that is, in (7), for $1 \leq q < \overline{q}$, we also have $\omega_q = (2\pi/T)(q - 1/2 - Q/2 + T)$.

In this section, we will exploit the first-order statistics (i.e., $E\{\mathbf{y}(n)\}$) of the received signal. (A consequence of using the first-order statistics is that the knowledge of the noise variance σ_v^2 in (H3) is not used here.)

By (H4), we have

$$c(n) = \sum_{m=0}^{P-1} c_m e^{j\alpha_m n} \quad \forall n, \qquad (8)$$

where

$$c_m := \frac{1}{P} \sum_{n=0}^{P-1} c(n) e^{-j\alpha_m n}, \quad \alpha_m := \frac{2\pi m}{P}. \qquad (9)$$

The coefficients c_m are known at the receiver since $\{c(n)\}$ is known. By (1)–(3), (8)-(9), and (H3), we have

$$E\{\mathbf{y}(n)\} = \sum_{q=1}^{Q} \sum_{m=0}^{P-1} \underbrace{\left[\sum_{l=0}^{L} c_m \mathbf{h}_q(l) e^{-j\alpha_m l} \right]}_{=:\mathbf{d}_{mq}} e^{j(\omega_q + \alpha_m)n} + \mathbf{m}. \qquad (10)$$

Suppose that we pick P to be such that $(\omega_q + \alpha_m)$'s are all distinct for any choice of m and q. For instance, suppose that the data record length T samples (see also Section 1.1) and P are such that $T = KP$ for some integer $K > 0$. In such a case, we have

$$\omega_{mq} \qquad\qquad\qquad\qquad\qquad\qquad (11)$$

$$:= \omega_q + \alpha_m$$

$$= \begin{cases} \dfrac{2\pi}{T}\left(q - \dfrac{1}{2} - \dfrac{Q}{2} + Km \right) \bmod(2\pi) & \text{if } Q \geq q \geq \dfrac{Q+1}{2}, \\[2mm] \dfrac{2\pi}{T}\left(q - \dfrac{1}{2} - \dfrac{Q}{2} + T + Km \right) \bmod(2\pi) & \text{if } 1 \leq q < \dfrac{Q+1}{2}. \end{cases} \qquad (12)$$

If P and K are such that $K \geq Q$, then it follows from (12) that $\omega_{m_1 q_1} \neq \omega_{m_2 q_2}$ if $q_1 \neq q_2$ or $m_1 \neq m_2$. Henceforth, it is assumed that the above conditions hold true. Then we have

$$T^{-1} \sum_{n=0}^{T-1} e^{j(2\pi/T)(q+Km)n} = \delta(q)\delta(m). \qquad (13)$$

Note that $\omega_{mq} = 0$ only when $m = 0$ and $q = \overline{q}$. We rewrite (10) as

$$E\{\mathbf{y}(n)\} = \sum_{q=1}^{Q} \sum_{\substack{m=0 \\ (q,m) \neq (\overline{q}, 0)}}^{P-1} \mathbf{d}_{mq} e^{j\omega_{mq} n} + (\mathbf{d}_{0\overline{q}} + \mathbf{m}). \qquad (14)$$

Given the observation sequence $\mathbf{y}(n)$, $0 \leq n \leq T - 1$, our approach to estimating $h_q(l)$'s using the first-order statistics of the data is to first estimate $\mathbf{d}_{mq}$'s for $0 \leq m \leq P - 1$, $1 \leq q \leq Q$ ($(q,m) \neq (\overline{q}, 0)$), and then estimate $h_q(l)$'s from the estimated $\mathbf{d}_{mq}$'s. By (14), $\mathbf{d}_{mq}$ is the coefficient of the exponential $e^{j\omega_{mq} n}$ for $(q,m) \neq (\overline{q}, 0)$, whereas $\mathbf{d}_{0\overline{q}} + \mathbf{m}$ is the coefficient of $e^{j\omega_{0\overline{q}} n} = 1$. Since the dc offset $\mathbf{m}$ is not necessarily known, we will not seek the coefficient of $e^{j\omega_{0\overline{q}} n}$ in (14). By (1)–(3) and (14), we have

$$\mathbf{y}(n) = \sum_{q=1}^{Q} \sum_{m=0}^{P-1} \left[\mathbf{d}_{mq} + \mathbf{m}\delta(q - \overline{q})\delta(m) \right] e^{j\omega_{mq} n} + \mathbf{e}(n), \qquad (15)$$

where $\mathbf{e}(n)$ is a zero-mean random sequence. Define the cost function

$$J = \sum_{n=0}^{T-1} \|\mathbf{e}(n)\|^2. \qquad (16)$$

Choose $\mathbf{d}_{mq}$'s ($q = 1, 2, \ldots, Q$; $m = 0, 1, \ldots, P-1, (q, m) \neq (\overline{q}, 0)$) to minimize J. For optimization, we must have

$$\frac{\partial J}{\partial \mathbf{d}_{mq}^*}\Bigg|_{\substack{\mathbf{d}_{mq} = \hat{\mathbf{d}}_{mq} \\ \forall q, m}} = 0, \qquad (17)$$

where the partial derivative in (17) for given m and q is a column vector of dimension N (the derivatives are componentwise). (17) leads to

$$\sum_{n=0}^{T-1} \mathbf{e}(n) e^{-j\omega_{mq} n}\Bigg|_{\substack{\mathbf{d}_{mq} = \hat{\mathbf{d}}_{mq} \\ \forall q, m}} = 0. \qquad (18)$$

Using (13), (15), and (18), it follows that (for $(q, m) \neq (\overline{q}, 0)$)

$$\hat{\mathbf{d}}_{mq} = \frac{1}{T} \sum_{n=0}^{T-1} \mathbf{y}(n) e^{-j\omega_{mq} n}. \qquad (19)$$

It follow from (14) and (19) that

$$E\{\hat{\mathbf{d}}_{mq}\} = \mathbf{d}_{mq}, \quad (q, m) \neq (\overline{q}, 0). \qquad (20)$$

Now we establish that given $\mathbf{d}_{mq}$ for $1 \leq q \leq Q$ and $0 \leq m \leq P-1$ but excluding $\omega_q + \alpha_m = 0$, we can (uniquely) estimate $\mathbf{h}_q(l)$'s if $P \geq L+2$ and $c_m \neq 0$ for all $m \neq 0$. Define

$$\mathbf{V} := \begin{bmatrix} 1 & e^{-j\alpha_1} & \cdots & e^{-j\alpha_1 L} \\ 1 & e^{-j\alpha_2} & \cdots & e^{-j\alpha_2 L} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\alpha_{P-1}} & \cdots & e^{-j\alpha_{P-1} L} \end{bmatrix}_{(P-1) \times (L+1)}, \qquad (21)$$

$$\mathbf{D}_m := [\mathbf{d}_{m1}^T, \mathbf{d}_{m2}^T, \ldots, \mathbf{d}_{mQ}^T]^T, \qquad (22)$$

$$\mathbf{H}_l := [\mathbf{h}_1^T(l), \mathbf{h}_2^T(l), \ldots, \mathbf{h}_Q^T(l)]^T, \qquad (23)$$

$$\mathcal{H} := \begin{bmatrix} \mathbf{H}_0^H & \mathbf{H}_1^H & \cdots & \mathbf{H}_L^H \end{bmatrix}^H, \qquad (24)$$

$$\mathcal{D}_1 := \begin{bmatrix} \mathbf{D}_1^H & \mathbf{D}_2^H & \cdots & \mathbf{D}_{P-1}^H \end{bmatrix}^H, \qquad (25)$$

$$\mathcal{C}_1 := \underbrace{(\text{diag}\{c_1, c_2, \ldots, c_{P-1}\} \mathbf{V})}_{=: \mathcal{V}} \otimes I_{NQ}. \qquad (26)$$

Omitting the term $m = 0$ and using the definition of $\mathbf{d}_{mq}$ from (10), it follows that

$$\mathcal{C}_1 \mathcal{H} = \mathcal{D}_1. \qquad (27)$$

Notice that we have omitted all pairs $(m, q) = (0, q)$ ($q \neq \overline{q}$) from (27). In order to include these omitted terms, we further define an $[N(Q-1)]$-column vector

$$\mathbf{D}_0 := [\mathbf{d}_{01}^T, \mathbf{d}_{02}^T, \ldots, \mathbf{d}_{0(\overline{q}-1)}^T \mathbf{d}_{0(\overline{q}+1)}^T, \ldots, \mathbf{d}_{0Q}^T]^T, \qquad (28)$$

an $[N(Q-1)] \times [NQ]$ matrix

$$\mathcal{A} := \begin{bmatrix} I_{N(\overline{q}-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{N(Q-\overline{q})} \end{bmatrix}, \qquad (29)$$

and an $[N(Q-1)] \times [NQ(L+1)]$ matrix

$$\mathcal{C}_2 := \begin{bmatrix} c_0 \mathcal{A} & c_0 \mathcal{A} & \cdots & c_0 \mathcal{A} \end{bmatrix}. \qquad (30)$$

Then it follows from (10) and (28)–(30) that

$$\mathcal{C}_2 \mathcal{H} = \mathbf{D}_0. \qquad (31)$$

In order to concatenate (27) and (31), we define

$$\mathcal{C} := \begin{bmatrix} \mathcal{C}_2 \\ \mathcal{C}_1 \end{bmatrix}, \qquad \mathcal{D} := \begin{bmatrix} \mathbf{D}_0 \\ \mathcal{D}_1 \end{bmatrix}, \qquad (32)$$

which lead to

$$\mathcal{C} \mathcal{H} = \mathcal{D}. \qquad (33)$$

Equation (33) utilizes all pairs (m, q) except $(0, \overline{q})$.

In (21) $\mathbf{V}$ is a Vandermonde matrix with a rank of $L+1$ if $P-1 \geq L+1$ and α_m's are distinct [14, page 274]. Since $c_m \neq 0$ for all m, by [14, Result R4, page 257], $\text{rank}(\mathcal{V}) = \text{rank}(\mathbf{V}) = L+1$. Finally, by [10, Property K6, page 431], $\text{rank}(\mathcal{C}_1) = \text{rank}(\mathcal{V}) \times \text{rank}(I_{NQ}) = NQ(L+1)$. Therefore, we can determine $\mathbf{h}_q(l)$'s uniquely from (27). Augmenting (27) with additional equations to obtain (33) keeps the earlier conclusions unchanged, that is, $\text{rank}(\mathcal{C}) = \text{rank}(\mathcal{C}_1) = NQ(L+1)$. Thus, if $P \geq L+2$ and $c_m \neq 0$ for all $m \neq 0$, (33) has a unique solution for $\mathcal{H}$ (i.e., $\mathbf{h}_q(l)$'s).

Define $\hat{\mathbf{D}}_m$ as in (22) or (28) with $\mathbf{d}_{mq}$'s replaced with $\hat{\mathbf{d}}_{mq}$'s. Similarly, define $\hat{\mathcal{D}}$ as in (25) and (32) with $\mathbf{D}_m$'s replaced with $\hat{\mathbf{D}}_m$'s. Then from (33) we have the channel estimate

$$\hat{\mathcal{H}} = (\mathcal{C}^H \mathcal{C})^{-1} \mathcal{C}^H \hat{\mathcal{D}}. \qquad (34)$$

By (20) and (33), it follows that

$$E\{\hat{\mathcal{H}}\} = \mathcal{H}. \qquad (35)$$

We summarize our method in the following lemma.

Lemma 1. *Under (H1)–(H4), the channel estimator (34) satisfies (35) under the following (additional) sufficient conditions: the periodic training sequence is such that $c_m \neq 0$ for all $m \neq 0$, $P \geq L+2$, and P and T are such that $T = KP$ for integer $K \geq Q$.*

Remark 1. A more logical approach would have been to select $\mathbf{h}_q(l)$'s and $\mathbf{m}$ jointly to minimize the cost J in (16). The resulting solution is more complicated and it couples estimates of $\mathbf{h}_q(l)$'s and $\mathbf{m}$. Since we do not use $\mathbf{d}_{0\overline{q}}$, we are discarding any information about $\mathbf{h}_{\overline{q}}(l)$ therein.

Remark 2. It should be emphasized that precise knowledge of the channel length L is not required; an upperbound L_u suffices. Then we estimate $\mathcal{H}_l$ for $0 \leq l \leq L_u$ with $E\{\hat{\mathcal{H}}_l\} = 0$ for $l \geq L+1$. Moreover, we do not need $c_m \neq 0$ for every m. We need at least $L+2$ nonzero c_ms.

Remark 3. The cost (16) is not novel; it also occurs in [1, 15] in the context of time-multiplexed training for doubly selective channels. However, unlike these papers, as noted in Remark 1 we do not directly estimate $\mathbf{h}_q(l)$'s and $\mathbf{m}$ (there is no $\mathbf{m}$ in these papers); rather, we first estimate $\mathbf{d}_{mq}$'s which are motivated through the time-varying mean $E\{\mathbf{y}(n)\}$, hence, the term first-order statistics. This aspect is missing from [1, 15], and in this paper it is motivated by the time-invariant results of [9, 16, 21] (among others). Choice of periodic superimposed training is also motivated by the results of [9, 16, 21].

3. DETERMINISTIC MAXIMUM-LIKELIHOOD (DML) APPROACH

The first-order statistics-based approach of Section 2 views the information sequence as interference. Since the training and information sequences of a given user pass through an identical channel, this fact can be exploited to enhance the channel estimation performance via an iterative approach. We now consider joint channel and information sequence estimation via an iterative DML (or conditional ML) approach assuming that the noise $\mathbf{v}(n)$ is complex Gaussian. We have guaranteed convergence to a local maximum. Furthermore, if we initialize with our superimposed training-based solution, one is guaranteed the global extremum (minimum error probability sequence estimator) if the superimposed training-based solution is "good."

Suppose that we have collected $T - L$ samples of the observations. Form the vector

$$Y = \left[\mathbf{y}^T(T-1), \mathbf{y}^T(T-2), \ldots, \mathbf{y}^T(L)\right]^T \quad (36)$$

and similarly define

$$\mathbf{s} := \left[s(T-1), s(T-2), \ldots, s(0)\right]^T. \quad (37)$$

Furthermore, let

$$\tilde{\mathbf{v}}(n) := \mathbf{v}(n) - \mathbf{m}. \quad (38)$$

Using (1)–(3) we then have the following linear model:

$$Y = \mathcal{T}(\mathbf{s})\mathcal{H} + \underbrace{\begin{bmatrix} \tilde{\mathbf{v}}(T-1) \\ \vdots \\ \tilde{\mathbf{v}}(L) \end{bmatrix}}_{=:\tilde{V}} + \underbrace{\begin{bmatrix} \mathbf{m} \\ \vdots \\ \mathbf{m} \end{bmatrix}}_{=:\mathcal{M}}, \quad (39)$$

where $V = \tilde{V} + \mathcal{M}$ is a column-vector consisting of samples of noise $\{\mathbf{v}(n)\}$ in a manner similar to (36), $\mathcal{H}$ is defined in (24), $\mathcal{T}(\mathbf{s})$ is a block Hankel matrix given by

$$\mathcal{T}(\mathbf{s}) := \begin{bmatrix} s(T-1)\Sigma_{T-1} & \cdots & s(T-L-1)\Sigma_{T-1} \\ s(T-2)\Sigma_{T-2} & \cdots & s(T-L-2)\Sigma_{T-2} \\ \vdots & \vdots & \vdots \\ s(L)\Sigma_L & \cdots & s(0)\Sigma_L \end{bmatrix}, \quad (40)$$

a block Hankel matrix has identical block entries on its block antidiagonals, and

$$\Sigma_n := \left[e^{j\omega_1 n}I_N \quad e^{j\omega_2 n}I_N \quad \cdots \quad e^{j\omega_Q n}I_N\right]. \quad (41)$$

Also using (1)–(3), an alternative linear model for Y is given by

$$Y = \mathcal{F}(\mathcal{H})\mathbf{s} + \tilde{V} + \mathcal{M}, \quad (42)$$

where

$$\mathcal{F}(\mathcal{H}) := \begin{bmatrix} \mathbf{h}(T-1;0) & \cdots & \mathbf{h}(T-1;L) & \\ & \ddots & & \ddots \\ & & \mathbf{h}(L;0) & \cdots & \mathbf{h}(L;L) \end{bmatrix} \quad (43)$$

is a "filtering matrix."

Consider (1), (3), and (39). Under the assumption of temporally white complex Gaussian measurement noise, consider the joint estimators

$$\{\widehat{\mathcal{H}}, \widehat{\mathbf{s}}, \widehat{\mathbf{m}}\} = \arg\left\{\min_{\mathcal{H}, \mathbf{s}, \mathbf{m}} \left\|Y - \mathcal{T}(\mathbf{s})\mathcal{H} - \mathcal{M}\right\|^2\right\}, \quad (44)$$

where $\widehat{\mathbf{s}}$ is the estimate of $\mathbf{s}$. In the above we have followed a DML approach assuming no statistical model for the input sequences $\{s(n)\}$. Using (39) and (42), we have a separable nonlinear least-squares problem that can be solved sequentially as (joint optimization with respect to $\mathcal{H}, \mathbf{m}$ can be further "separated")

$$\begin{aligned} \{\widehat{\mathcal{H}}, \widehat{\mathbf{s}}, \widehat{\mathbf{m}}\} &= \arg\min_{\mathbf{s}}\left\{\min_{\mathcal{H}, \mathbf{m}}\left\|Y - \mathcal{T}(\mathbf{s})\mathcal{H} - \mathcal{M}\right\|^2\right\} \\ &= \arg\min_{\mathcal{H}, \mathbf{m}}\left\{\min_{\mathbf{s}}\left\|Y - \mathcal{F}(\mathcal{H})\mathbf{s} - \mathcal{M}\right\|^2\right\}. \end{aligned} \quad (45)$$

The finite alphabet properties of the information sequences can also be incorporated into the DML methods. These algorithms, first proposed by Seshadri [13] for time-invariant SISO systems, iterate between estimates of the channel and the input sequences. At iteration k, with an initial guess of the channel $\mathcal{H}^{(k)}$ and the mean $\mathbf{m}^{(k)}$, the algorithm estimates the input sequence $\mathbf{s}^{(k)}$ and the channel $\mathcal{H}^{(k+1)}$ and mean $\mathbf{m}^{(k+1)}$ for the next iteration by

$$\mathbf{s}^{(k)} = \arg\min_{\mathbf{s} \in \mathscr{S}}\left\|Y - \mathcal{F}(\mathcal{H}^{(k)})\mathbf{s} - \mathcal{M}^{(k)}\right\|^2, \quad (46)$$

$$\mathcal{H}^{(k+1)} = \arg\min_{\mathcal{H}}\left\|Y - \mathcal{T}(\mathbf{s}^{(k)})\mathcal{H} - \mathcal{M}^{(k)}\right\|^2, \quad (47)$$

$$\mathbf{m}^{(k+1)} = \arg\min_{\mathbf{m}}\left\|Y - \mathcal{T}(\mathbf{s}^{(k)})\mathcal{H}^{(k+1)} - \mathcal{M}\right\|^2, \quad (48)$$

where $\mathscr{S}$ is the (discrete) domain of $\mathbf{s}$. The optimizations in (47) and (48) are linear least squares problems whereas the the optimization in (46) can be achieved by using the Viterbi algorithm [11]. Note that (46)–(48) can be interpreted as a constrained alternating least-squares implementation with $\mathbf{s} \in \mathscr{S}$ as the constraint. Since the above iterative procedure involving (46), (47), and (48) decreases the cost at every iteration, one achieves a local maximum of the DML function.

We now summarize our DML approach in the following steps.

(1)   (a) Use (34) to estimate the channel using the first-order (cyclostationary) statistics of the observations. Denote the channel estimates by $\widehat{\mathcal{H}}^{(1)}$ and $\hat{\mathbf{h}}_q^{(1)}(l)$. In this method $\{c(n)\}$ is known and $\{b(n)\}$ is regarded as interference.

(b) Estimate the mean $\hat{\mathbf{m}}^{(1)}$ as follows. Define (recall (1)–(3))

$$\hat{\mathbf{m}}^{(1)} := \left(\frac{1}{T}\right)\sum_{n=0}^{T-1}\left[\mathbf{y}(n) - \sum_{l=0}^{L}\hat{\mathbf{h}}^{(1)}(n;l)c(n-l)\right],$$
$$\hat{\mathbf{h}}^{(1)}(n;l) := \sum_{q=1}^{Q}\hat{\mathbf{h}}_q^{(1)}(l)e^{j\omega_q n}. \tag{49}$$

(c) Design a Viterbi sequence detector to estimate $\{s(n)\}$ as $\{\tilde{s}(n)\}$ using the estimated channel $\widehat{\mathcal{H}}^{(1)}$, mean $\hat{\mathbf{m}}^{(1)}$ and cost (46) with $k = 1$. (Note that knowledge of $\{c(n)\}$ is used in $s(n) = b(n) + c(n)$, therefore, we are in essence estimating $b(n)$ in the Viterbi detector.)

(2)   (a) Substitute $\tilde{s}(n)$ for $s(n)$ in (1) and use the corresponding formulation in (39) to estimate the channel $\mathcal{H}$ as

$$\widehat{\mathcal{H}}^{(2)} = \mathcal{T}^{\dagger}(\tilde{\mathbf{s}})[Y - \widehat{\mathcal{M}}^{(1)}]. \tag{50}$$

Define $\hat{\mathbf{h}}^{(2)}(n;l)$ using $\hat{\mathbf{h}}_q^{(2)}(l)$ in a manner similar to $\hat{\mathbf{h}}^{(1)}(n;l)$. Then the mean $\mathbf{m}$ is estimated as $\hat{\mathbf{m}}^{(2)}$ given by

$$\mathbf{m}^{(2)} = \frac{1}{T-L}\sum_{n=L}^{T-1}\left[\mathbf{y}(n) - \sum_{l=0}^{L}s^{(1)}(n-l)\mathbf{h}^{(2)}(n;l)\right]. \tag{51}$$

(b) Design a Viterbi sequence detector using the estimated channel $\widehat{\mathcal{H}}^{(2)}$, mean $\hat{\mathbf{m}}^{(2)}$, and cost (46) with $k = 2$, as in step (1)(c).

(3) Step (2) provides one iteration of (46)-(47). Repeat a few times till any (relative) improvement in channel estimation over previous iteration is below a pre-specified threshold.

4.   SIMULATION EXAMPLES

We now present several computer simulation examples in support of our proposed approach. Example 1 uses an exact CE-BEM representation to generate data whereas Examples 2 and 3 use a 3-tap Jakes' channel to generate data. In all examples CE-BEMs are used to process the observations; therefore, in Examples 2 and 3 we have approximate modeling.

Example 1. In this example we pick an arbitrary value of Q independent of T. In (2) take $N = 1$, $Q = 2$, and

$$\omega_1 = 0, \qquad \omega_2 = \frac{2\pi}{50}. \tag{52}$$

We consider a randomly generated channel in each Monte Carlo run with random channel length $L \in \{0, 1, 2\}$ picked with equal probabilities and random channel coefficients $h_q(l)$, $0 \le l \le L$, taken to be mutually independent complex random variables with independent real and imaginary parts, each uniformly distributed over the interval $[-1, 1]$. Normalized mean-square error (MSE) in estimating the channel coefficients $h_q(l)$, averaged over 100 Monte Carlo runs, was taken as the performance measure for channel identification. It is defined as (before Monte Carlo averaging)

$$\text{NCMSE}_1 := \frac{\left\{\sum_{q=1}^{Q}\sum_{m=0}^{2}||h_q(m) - \hat{h}_q(m)||^2\right\}}{\sum_{q=1}^{Q}\sum_{m=0}^{2}||h_q(m)||^2} \tag{53}$$

The training sequence was taken to be an m-sequence (maximal length pseudorandom binary sequence) of length $7 (= P)$

$$\{c(n)\}_{n=0}^{6} = \{1, -1, -1, 1, 1, 1, -1\}. \tag{54}$$

The input information sequence $\{b(n)\}$ is i.i.d. equiprobable 4-QAM. As in [9, 16], define a power loss factor

$$\alpha = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2} \tag{55}$$

and power loss $-10\log(\alpha)$ dB, as a measure of the information data power loss due to the inclusion of the training sequence. Here

$$\sigma_b^2 := E\left\{|b(n)|^2\right\}, \qquad \sigma_c^2 := \frac{1}{P}\sum_{n=0}^{P-1}|c(n)|^2. \tag{56}$$

The training sequence was scaled to achieve a desired power loss. Complex white zero-mean Gaussian noise was added to the received signal and scaled to achieve a desired signal-to-noise (SNR) ratio at the receiver (relative to the contribution of $\{s(n)\}$).

Our proposed method using $L = L_u = 4$ (channel length overfit) in (34) was applied for varying power losses due to the superimposed training sequence. Figure 1 shows the simulation results. It is seen that as α decreases (i.e., training power increases relative to the information sequence power), one gets better results. Moreover, the proposed method works with overfitting. Finally, adding nonzero mean (dc offset) to additive noise yielded essentially identical results (differences do not show on the plotted curves).

Example 2. Consider (1) with $N = 1$ and $L = 2$. We simulate a random time-and frequency-selective Rayleigh fading channel following [20]. For different l's, $h(n;l)$'s are mutually independent and for a given l, we follow the modified Jakes' model [20] to generate $h(n;l)$:

$$h(n;l) = X(t)|_{t=nT_s}, \tag{57}$$

where $X(t) = (2/\sqrt{M})\sum_{i=1}^{M}e^{j\psi_i}\cos(2\pi f_d t\cos(\alpha_i) + \phi)$, $\alpha_i = (2\pi i - \pi + \theta)/(4M)$, $i = 1, 2, \ldots, M$, random variables θ, ϕ, and ψ_i are mutually independent ($\forall i$) and uniformly distributed
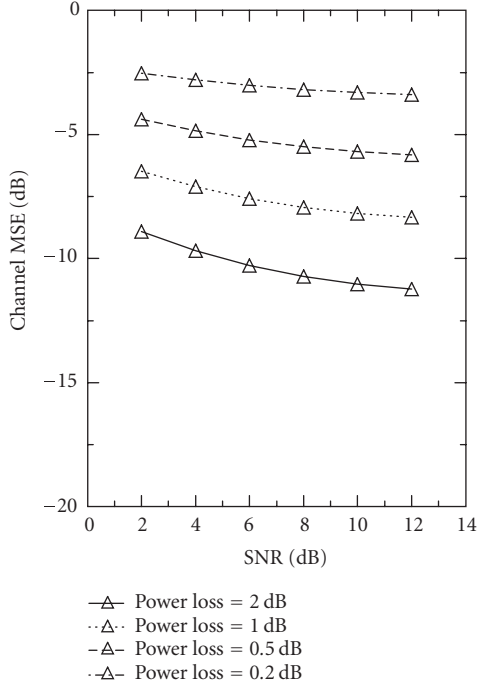
FIGURE 1: Example 1. Normalized channel MSE (53) based on $T = 140$ symbols per run, 100 Monte Carlo runs, QPSK signal, $P = 7$. Power loss $= -10 \log(\alpha)$ dB where α is as in (55).

over $[0, 2\pi)$, T_s denotes the symbol interval, f_d denotes the (max.) Doppler spread, and $M = 25$. For a fixed l, (57) generates a random process $\{h(n;l)\}_n$ whose power spectrum approximates the Jakes' spectrum as $M \uparrow \infty$. We consider a system with carrier frequency of 2 GHz, data rate of 40 kB (kB = kilo-Bauds), therefore, $T_s = 25 \times 10^{-6}$ seconds, and a varying Doppler spread f_d in the range 0 Hz to 200 Hz (corresponding to a maximum mobile velocity in the range 0 to 108 km/hr). We picked a data record length of 400 symbols (time duration of 10 msec). For a given Doppler spread, we pick Q as in Section 1.1 ($T = 400$, $L = 2$ in (7)). For the chosen parameters it varies within the values $\{1, 3, 5\}$. We emphasize that the CE-BEM was used only for processing at the receiver; the data were generated using (57).

We take all sequences (information and training) to be binary. For superimposed training, we take a periodic (scaled) binary sequence of period $P = 7$ with the training-to-information sequence power ratio (TIR) of 0.3 where

$$\text{TIR} = \frac{\sigma_c^2}{\sigma_b^2} = \alpha^{-1} - 1 \qquad (58)$$

and σ_b^2 and σ_c^2 denote the average power in the information sequence $\{b(n)\}$ and training sequence $\{c(n)\}$, respectively. Complex white zero-mean Gaussian noise was added to the received signal and scaled to achieve a target bit SNR at the receiver (relative to the contribution of $\{s(n)\}$).

For comparison, we consider conventional time-multiplexed training assuming time-invariant channels, as well as

CE-BEM-based periodically placed time-multiplexed training with and without zero-padding, following [7]. In the former, the block of data of length 400 symbols was split into two nonoverlapping blocks of 200 symbols each. Each subblock had a training sequence length of 46 symbols in the middle of the data subblock with 154 symbols for information; this leads to a training-to-information sequence power ratio (over the block length) of approximately 0.3. Assuming synchronization, time-invariant channels were estimated using conventional training and used for information detection via a Viterbi algorithm; this was done for each subblock. In the CE-BEM set-up, following [7], we took a training block of length $2L + 1 = 5$ and a data block of length 17 bits leading to a frame of length 22 bits. This frame was repeated over the entire record length (22×18). Thus, we have a training-to-information bit ratio of approximately 0.3. Two versions of training sequences were considered. In one of them zero-padding was used with a random bit in the middle of the training block, as in [7]: this leads to a peak-to-average power ratio (PAR) of 5. In the other version we had a random binary sequence of length 5 in each training block, leading to a PAR of 1 (an ideal choice). Assuming synchronization, CE-BEM channel was estimated using conventional training and used for information detection via a Viterbi algorithm. We also considered another variation of zero-padded training with a training block of length $2L + 1 = 5$ but a data block of length 50 bits leading to a training-to-information bit ratio of 0.1. Thus the proposed superimposed training scheme results in a data transmission rate that is 30% higher than the data transmission rate in all of the time-multiplexed training schemes considered in this example, except for the last scheme compared to which the data transmission rate is 10% higher.

Figure 2 shows the BER (bit error rate) based on 500 Monte Carlo runs for conventional training based on time-invariant (TI) modeling, the CE-BEM-based periodically placed time-multiplexed training for PAR = 5 and PAR = 1, the first-order statistics and superimposed training-based method and the proposed DML approach with two iterations, under varying Doppler spreads f_d and a bit SNR of 25 dB. It is seen that as Doppler spread f_d increases beyond about 60 Hz (normalized Doppler $T_s f_d$ of 0.0015), superimposed training approach of Section 2 (step (1)) outperforms the conventional (midamble) training with time-invariant channel approximation, without decreasing the data transmission rate. Furthermore, the proposed DML enhancement can lead to a significant improvement with just one iteration. On the other hand, the CE-BEM-based periodically placed time-multiplexed training approach of [7] significantly outperforms the superimposed training-based approaches, but at the cost of a reduction in the data transmission rate. Figure 3 shows the normalized channel mean-square error (NCMSE), defined (before averaging over runs) as

$$\text{NCMSE} = \frac{\sum_{n=1}^{T} \sum_{l=0}^{2} |\hat{h}(n;l) - h(n;l)|^2}{\sum_{n=1}^{T} \sum_{l=0}^{2} |h(n;l)|^2}. \qquad (59)$$

It is seen that the proposed DML enhancement leads to a significant improvement in channel estimation also with just one iteration.
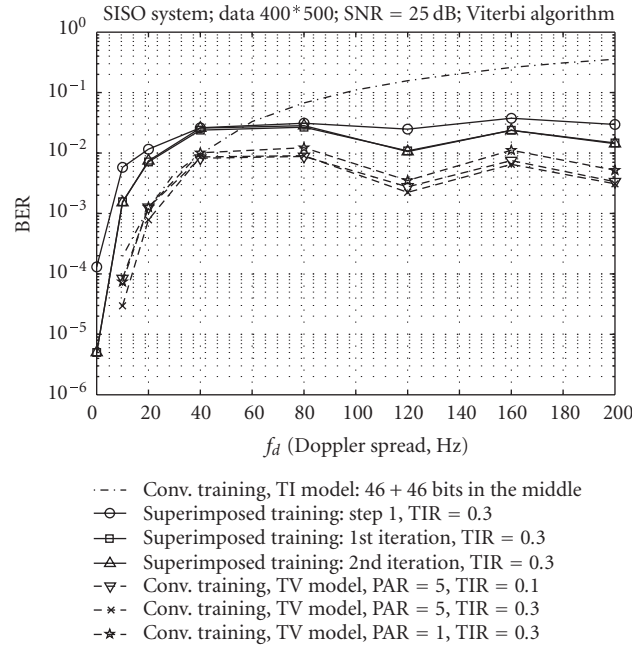
SISO system; data 400*500; SNR = 25 dB; Viterbi algorithm

- ·-·- Conv. training, TI model: 46 + 46 bits in the middle
- —○— Superimposed training: step 1, TIR = 0.3
- —□— Superimposed training: 1st iteration, TIR = 0.3
- —△— Superimposed training: 2nd iteration, TIR = 0.3
- -▽- Conv. training, TV model, PAR = 5, TIR = 0.1
- -✕- Conv. training, TV model, PAR = 5, TIR = 0.3
- -★- Conv. training, TV model, PAR = 1, TIR = 0.3

FIGURE 2: Example 2. BER: circle: estimate channel using superimposed training (training-to-information symbol power ratio TIR = 0.3) and then design a Viterbi detector; square: first iteration specified by step (2) (Section 3); up-triangle: second iteration specified by step (2) (Section 3); dot-dashed: estimate channel using conventional time-multiplexed training of length 46 bits in the middle of a subblock of length 200 bits and then design a Viterbi detector; cross: CE-BEM-based periodically placed time-multiplexed training with zero padding [7], TIR = 0.3; star: CE-BEM-based periodically placed time-multiplexed training without zero padding, TIR = 0.3; down-triangle: CE-BEM-based periodically placed time-multiplexed training with zero-padding [7], TIR = 0.1. SNR = 25 dB. Record length = 400 bits. Results are based on 500 Monte Carlo runs.
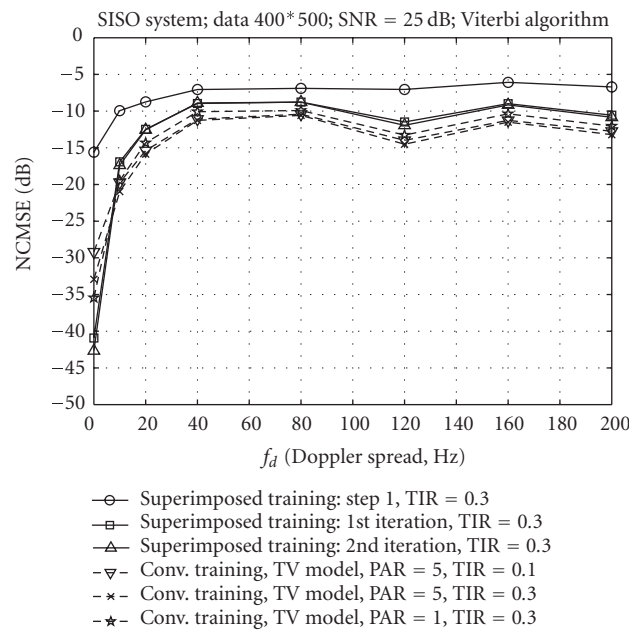

SISO system; data 400*500; SNR = 25 dB; Viterbi algorithm

- —○— Superimposed training: step 1, TIR = 0.3
- —□— Superimposed training: 1st iteration, TIR = 0.3
- —△— Superimposed training: 2nd iteration, TIR = 0.3
- -▽- Conv. training, TV model, PAR = 5, TIR = 0.1
- -✕- Conv. training, TV model, PAR = 5, TIR = 0.3
- -★- Conv. training, TV model, PAR = 1, TIR = 0.3

FIGURE 3: Example 2. As in Figure 2 except that NCMSE (normalized channel mean-square error) (59) is shown.

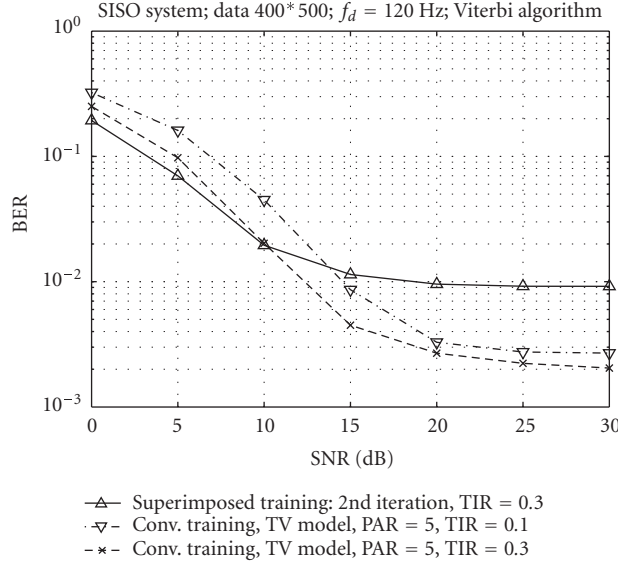SISO system; data 400*500; f_d = 120 Hz; Viterbi algorithm

FIGURE 4: Example 3. BER for varying SNR with Doppler spread f_d = 120 Hz: up-triangle: superimposed training, second iteration specified by step (2) (Section 3), TIR = 0.3; cross: CE-BEM-based periodically placed time-multiplexed training with zero padding [7], TIR = 0.3; down-triangle: CE-BEM-based periodically placed time-multiplexed training with zero padding [7], TIR = 0.1. After estimating the channel, we design a Viterbi detector using the estimated channel. Record length = 400 bits. Results are based on 500 Monte Carlo runs.
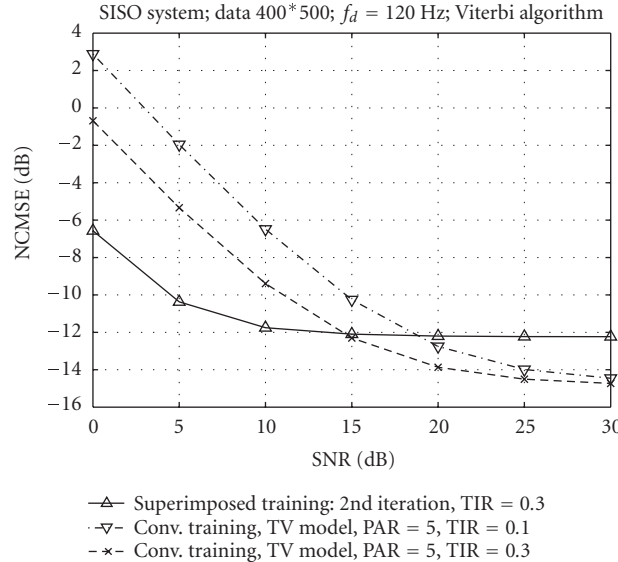
SISO system; data 400*500; f_d = 120 Hz; Viterbi algorithm

FIGURE 5: Example 3. As in Figure 4 except that corresponding NCMSE (normalized channel mean-suare error) (59) is shown.

Example 3. To further compare the relative advantages and disadvantages of CE-BEM-based superimposed training and periodically placed time-multiplexed training, we now repeat Example 2 but with varying SNR; the other details remain unchanged. Figures 4 and 5 show the simulation results for a Doppler spread of 120 Hz (normalized Doppler spread of 0.003 for bit duration of T_s = 25 μs) where we compare the results of the second iteration of the proposed DML approach based on superimposed training with that of periodically placed time-multiplexed training. There is an error floor with increasing SNR which is attributable to modeling

errors in approximating the Jakes' model with CE-BEM. It is seen from Figure 4 that our proposed approach outperforms (better BER) the CE-BEM-based periodically placed time-multiplexed training approach of [7] for SNRs at or below 10 dB, and underperforms for SNRs at or above 20 dB. There is also the data transmission rate advantage at all SNRs.

5. CONCLUSIONS

In this paper we first presented and extended the first-order statistics-based approach of [17] for time-varying (CE-BEM

based) channel estimation using superimposed training. Then we extended the first-order statistics-based solution to an iterative approach to joint channel and information sequence estimation, based on CE-BEM, using Viterbi detectors. The first-order statistics-based approach views the information sequence as interference whereas in the iterative joint estimation version it is exploited to enhance channel estimation and information sequence detection. The results were illustrated via several simulation examples some of them involving time-and frequency-selective Rayleigh fading where we compared the proposed approaches to some of the existing approaches. Compared to the CE-BEM-based periodically placed time-multiplexed training approach of [7], one achieves a lower BER for SNRs at or below 10 dB, and higher BER for SNRs at or above 20 dB. There is also the data transmission rate advantage at all SNRs. Further work is needed to compare the relative advantages and disadvantages of CE-BEM-based superimposed training and periodically placed time-multiplexed training.
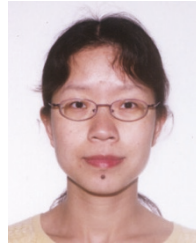
ACKNOWLEDGMENTS

REFERENCES

[1] M.-A. R. Baissas and A. M. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 2037–2049, 2002.

[2] I. Barhumi, G. Leus, and M. Moonen, "Time-varying FIR equalization for doubly selective channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 202–214, 2005.

[3] B. Farhang-Boroujeny, "Pilot-based channel identification: proposal for semi-blind identification of communication channels," *Electronics Letters*, vol. 31, no. 13, pp. 1044–1046, 1995.

[4] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1969–1986, 1998.

[5] G. Leus, "Semi-blind channel estimation for rapidly time-varying channels," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 3, pp. 773–776, Philadelphia, Pa, USA, March 2005.

[6] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1832–1840, 2003.

[7] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1351–1366, 2003.

[8] X. Meng and J. K. Tugnait, "Superimposed training-based doubly-selective channel estimation using exponential and polynomial bases models," in *Proceedings of the 38th Annual Conference on Information Sciences & Systems (CISS '04)*, Princeton University, Princeton, NJ, USA, March 2004.

[9] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel estimation using implicit training," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 240–254, 2004.

[10] B. Porat, *Digital Processing of Random Signals*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1994.

[11] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.

[12] P. Schniter, "Low-complexity equalization of OFDM in doubly selective channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1002–1011, 2004.

[13] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Transactions on Communications*, vol. 42, no. 2–4, part 2, pp. 1000–1011, 1994.

[14] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1997.

[15] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *International Journal of Adaptive Control & Signal Processing*, vol. 10, no. 2-3, pp. 159–176, 1996.

[16] J. K. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Communications Letters*, vol. 7, no. 9, pp. 413–415, 2003.

[17] _____, "On channel estimation using superimposed training and first-order statistics," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '03)*, vol. 4, pp. 624–627, Hong Kong, April 2003.

[18] J. K. Tugnait and X. Meng, "On superimposed training for channel estimation: performance analysis, training power allocation, and frame synchronization," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 752–765, 2006.

[19] J. K. Tugnait, L. Tong, and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 16–28, 2000.

[20] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 920–928, 2003.

[21] G. T. Zhou, M. Viberg, and T. McKelvey, "A first-order statistical method for channel estimation," *IEEE Signal Processing Letters*, vol. 10, no. 3, pp. 57–60, 2003.

Jitendra K. Tugnait received the B.S. degree with honors in electronics and electrical communication engineering from the Punjab Engineering College, Chandigarh, India, in 1971, the M.S. and the E.E. degrees from Syracuse University, Syracuse, NY, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, in 1973, 1974, and 1978, respectively, all in electrical engineering. From 1978 to 1982, he was an Assistant Professor of electrical and computer engineering at the University of Iowa, Iowa City, Iowa. He was with the Long Range Research Division of the Exxon Production Research Company, Houston, Tex, from June 1982 to September 1989. He joined the Department of Electrical and Computer Engineering, Auburn University, Auburn, Aa, in September 1989 as a Professor. He currently holds the title of James B. Davis Professor. His current research interests are in statistical signal processing, wireless and wireline digital communications, multiple-sensor multiple-target tracking, and stochastic systems analysis. Dr. Tugnait is a past Associate Editor of the IEEE Transactions on Automatic Control and of the IEEE Transactions

on Signal Processing. He is currently an Editor of the IEEE Transactions on Wireless Communications and an Associate Editor of IEEE Signal Processing Letters. He was elected Fellow of IEEE in 1994.

Xiaohong Meng was born in Luoyang, He'nan Province, China, on June 12, 1973. She received her B.E. degree in 1995 in electrical engineering from Beijing University of Posts and Telecommunications. From 1995 to 1999, she held the position of Instructor at He'nan Posts and Telecommunications School. From September 1995 to June 2001, she studied as a graduate student at Beijing University of Posts and Telecommunications. From January 2002 to May 2005, she was a Research Assistant in Electrical and Computer Engineering Department of Auburn University. She received her Ph.D. degree in electrical engineering in May 2005 and her M.S. degree in mathematics in May 2006 from Auburn University. She joined MIPS Technologies, Mountain View, Calif, in March 2006. Her current research interests include digital signal processing, statistical signal processing, wireless communications, and semiblind equalization.

Shuangchi He received the B.E. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2000 and 2003, respectively. He is currently working towards his Ph.D. degree in electrical engineering at Auburn University. Since 2003, he has been a Graduate Research Assistant at the Department of Electrical and Computer Engineering, Auburn University. His research interests include channel estimation and equalization, multiuser detection, and statistical and adaptive signal processing and analysis.