14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

ESTIMATION OF DOUBLY-SELECTIVE CHANNELS IN BLOCK TRANSMISSIONS USING DATA-DEPENDENT SUPERIMPOSED TRAINING

Mounir Ghogho[1] *and Ananthram Swami*[2]
[1] *School of EEE, Univ. of Leeds, LS29JT, United Kingdom*
[2] *Army Research Laboratory, Adelphi, MD 20783-1197, USA*

ABSTRACT

We propose to estimate time-varying frequency-selective channels using data-dependent superimposed training (DDST) and a basis expansion model (BEM). The proposed method is an extension of the DDST-based method recently proposed for time-invariant channels. The superimposed training consists of the sum of a known sequence and a data-dependent sequence, which is unknown to the receiver. The data-dependent sequence cancels the effects of the unknown data on channel estimation. Simulation results show that the proposed method compares favorably with time-division multiplexing training.

1. INTRODUCTION

Wireless and mobile communications channels for high data rate transmission are typically time and frequency selective. Frequency-selectivity is due to multipath propagation and large signal bandwidth whereas time-selectivity is induced by Doppler. Such doubly selective channels offer joint multipath-Doppler diversity gains [1, 2]. However, achieving such gains requires channel acquisition, which is a challenging task. Further, when the channel is fast fading, the common approach of assuming the channel to be quasi-static over a certain interval of time may lead to unacceptable system performance. Thus, accurate estimation of doubly-selective channels is well motivated.

In most practical systems, training is used to facilitate channel estimation. Blind techniques typically require long data records and are often complex to implement. The conventional way of multiplexing training symbols with the data is time-division multiplexing (TDM) [3, 4]. In the case of purely time-selective channels, periodic insertion of training symbols, known as pilot symbol aided modulation (PSAM), was shown to be optimal in the sense of minimizing the mean square error (MSE) of channel estimation. For purely frequency-selective channels, periodic insertion of pilot clusters was shown to be optimal [4]. For doubly selective channels and zero-padded block transmission, using the basis expansion model (BEM) of [5], it was shown in [6] that periodic insertion of zero-guarded pilot symbols was optimal. For cyclic-prefixed systems, orthogonal multiplexing is implemented in the frequency domain [7]. An alternative approach to orthogonal multiplexing schemes is superimposed training (ST). This scheme saves valuable bandwidth at the expense of a reduction in the information signal-to-noise ratio (SNR), since some of the transmitted energy is allocated to the embedded pilots. In the case of purely time-selective channels, it was shown in [8] that ST outperforms PSAM when the fading is fast. ST schemes have also been proposed in [9, 10, 11]. The main drawback of such a scheme is that performance of a channel estimator is limited by the unknown data which act as a source of input noise. To circumvent this, a variant of the ST scheme, called data-dependent ST (DDST) was proposed in [12, 13] for purely frequency-selective channels. Unlike the conventional ST scheme, the training sequence in the DDST method was set to be the sum of a known (to the receiver) sequence, and a *data-dependent* sequence, which is unknown at the receiver. Here, we extend this method to include time and frequency-selective channels. Towards this objective, we use the basis expansion model (BEM) [5, 1] which has been used to approximate doubly selective channels.

The paper is organized as follows. The next section describes the system model. Channel estimation is presented in Section 3. The issue of optimum training design is addressed for the proposed pilot assisted transmission in Section 4. Equalization and symbol detection are explained in Section 5. Simulations results are presented in Section 6 and conclusions are drawn in Section 7.

Notation: Superscripts *, ^{T}, ^{H} and † denote complex conjugate, transpose, Hermitian and pseudo-inverse operators respectively. The trace and statistical expectation are denoted by $\mathrm{Tr}\{\cdot\}$ and $E\{\cdot\}$. The nth element of a vector $\boldsymbol{z}$ is denoted by $z(n)$. The $(N \times N)$ identity matrix is denoted by $\mathbf{I}$. Finally, $\mathrm{diag}(a_1, ..., a_N)$ is the $(N \times N)$ diagonal matrix whose nth diagonal entry is a_n. A matrix of zero will be denoted by $\mathbf{0}$. The symbol $\propto$ will mean "proportional to" and $< . >_N$ denotes arithmetic modulo N.

2. SYSTEM MODEL

Consider a doubly-selective communication link and let $h(t;\tau)$ denote its time-varying impulse response which includes the doubly-selective channel as well as the transmit-receive filters. Let $H(f;\tau)$ denote the Fourier transform of $h(t;\tau)$. Let us define the delay spread $\tau_{\max}$ and the Doppler spread $f_{\max}$ as the thresholds on τ and f beyond which $|H(f;\tau)| \approx 0$.

Consider a cyclic-prefixed single-carrier block transmission system operating over such a channel. In order to avoid interblock interference, we assume the length of the cyclic prefix (CP) to be larger or equal to the length of the channel. At the receiver, after removing the CP, the baud-sampled discrete-time baseband signal model for each received block (we omit the block index for notational simplicity) is

$$y(n) = \sum_{\ell=0}^{L-1} h(n;\ell)s(n-\ell) + v(n), \qquad n = 0 \cdots N-1 \quad (1)$$

where N is the length of the block, $h(n;\ell)$ is the time-varying ℓth tap of the channel, $L-1$ is the order of the channel in number of samples, and $\{s(n)\}$ is the transmitted block. Because of the CP $s(-i) = s(N-i)$, $i = 1 \cdots L-1$. We assume that the transmitted symbols $s(n)$ are zero-mean and independent of the zero-mean noise $v(n)$.

Now, we use a BEM to model the time-varying channel. We focus here on the exponential basis functions. Under the assumption that the delay and Doppler spreads are bounded by τ_{max} and f_{max} respectively, the time-varying channel can be modelled for $n = 0, \cdots, N-1$ as

$$h(n;\ell) = \sum_{q=-Q/2}^{Q/2} h_{q,\ell} e^{j2\pi qn/N}, \qquad \ell = 0, ..., L-1 \tag{2}$$

where L and Q satisfy the following conditions:

$$(L-1)T \geq \tau_{max} \qquad Q/(NT) \geq 2f_{max}$$

where T is the symbol period. We also assume that $N >> L(Q+1)$.

Using eq. (2), the signal model in eq. (1) can be written in vector form as

$$\boldsymbol{y} = \sum_{q=-Q/2}^{Q/2} \mathbf{D}_q \mathcal{H}_q \boldsymbol{s} + \boldsymbol{v} \tag{3}$$

where $\mathbf{D}_q := \text{diag}\left(1, \cdots, e^{j2\pi q(N-1)/N}\right)$, $\mathcal{H}_q$ is an $(N \times N)$ circular matrix with first column given by $[h_{q,0}, h_{q,1}, ...h_{q,L-1}, 0, ..., 0]^T$, and $\boldsymbol{s}$ is the $(N \times 1)$ transmitted block. Equivalently, $\boldsymbol{y}$ can be rewritten as

$$\boldsymbol{y} = \sum_{q=-Q/2}^{Q/2} \mathbf{D}_q \mathcal{S} \boldsymbol{h}_q + \boldsymbol{v} \tag{4}$$

where $\mathcal{S}$ is the leading $(N \times L)$ matrix of the $(N \times N)$ circular Toeplitz matrix whose first column is $\boldsymbol{s}$, and $\boldsymbol{h}_q = [h_{q,0}, ..., h_{q,L-1}]^T$. In a more compact form, $\boldsymbol{y}$ can be expressed as

$$\boldsymbol{y} = \mathbf{D}[\mathbf{I}_{Q+1} \otimes \mathcal{S}]\boldsymbol{h} + \boldsymbol{v} \tag{5}$$

where $\mathbf{D} = [\mathbf{D}_{-Q/2} \cdots \mathbf{D}_{Q/2}]$, $\boldsymbol{h} = [\boldsymbol{h}_{-Q/2}^T \cdots \boldsymbol{h}_{Q/2}^T]^T$.

3. CHANNEL ESTIMATION USING DDST

In a TDM scheme, some of the entries of $\boldsymbol{s}$ are known pilots. In the conventional ST scheme, a known training sequence, $\boldsymbol{c}$, is added to the data vector, $\boldsymbol{w}$, i.e., $\boldsymbol{s} = \boldsymbol{w} + \boldsymbol{c}$. The data symbols are assumed to be zero-mean, i.i.d. random variables drawn from a finite alphabet, e.g., PSK or QAM; let σ_w^2 denote the data symbol power. The channel coefficients can be consistently estimated using the first-order statistics of the received signal [9, 10]. A disadvantage of this method is that the channel estimate is degraded by the embedded unknown data, which acts as a source of input noise. To mitigate this problem, we use the DDST approach [12], where $\boldsymbol{w}$ is distorted prior to adding the known training sequence. Let $\underline{\boldsymbol{w}} = \boldsymbol{w} + \boldsymbol{e}$ be the distorted data vector, where $\boldsymbol{e}$ is a zero-mean data-dependent sequence. With $\boldsymbol{s} = \underline{\boldsymbol{w}} + \boldsymbol{c}$, $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \mathbf{D}[\mathbf{I}_{Q+1} \otimes \boldsymbol{\mathcal{C}}]\boldsymbol{h} + \mathbf{D}[\mathbf{I}_{Q+1} \otimes \underline{\mathcal{W}}]\boldsymbol{h} + \boldsymbol{v}$$

where $\boldsymbol{\mathcal{C}}$ and $\underline{\mathcal{W}}$ are defined similar to $\mathcal{S}$ in (4). The linear least squares (LLS) channel estimate, which regards the data-related term on the RHS of the above equation as noise, is given by

$$\hat{\boldsymbol{h}} = \left(\mathbf{D}[\mathbf{I}_{Q+1} \otimes \boldsymbol{\mathcal{C}}]\right)^{\dagger} \boldsymbol{y}. \tag{6}$$

Note that if the statistics of the channel are available, the minimum mean square error (MMSE) estimator could be used. Here, to simplify the equations, we use the above LLS estimator.

3.1. Identifiability

The channel estimator in eq. (6) is consistent iff the following identifiability condition is satisfied

$$\text{rank}\left\{\mathbf{D}[\mathbf{I}_{Q+1} \otimes \boldsymbol{\mathcal{C}}]\right\} = L(Q+1). \tag{7}$$

Equivalently,

$$\text{rank}\left\{[\mathbf{D}_{-Q/2}\boldsymbol{\mathcal{C}}, ..., \mathbf{D}_0\boldsymbol{\mathcal{C}}, ..., \mathbf{D}_{Q/2}\boldsymbol{\mathcal{C}}]\right\} = L(Q+1).$$

Recall that $\boldsymbol{\mathcal{C}}$ is the $N \times L$ leading submatrix of a circulant matrix, and the $\mathbf{D}$'s by definition are $N \times N$ full rank matrices. Equivalently $N \geq L(Q+1)$, $\boldsymbol{\mathcal{C}}$ must have full column rank L, and the $Q+1$ sub-matrices must be orthogonal to one-another, i.e., $\boldsymbol{\mathcal{C}}^H \mathbf{D}_m^H \mathbf{D}_n \boldsymbol{\mathcal{C}} = \mathbf{0}$, $m, n = -Q/2, ..., Q/2$, $m \neq n$. We note that such a condition would be required even if the exponential bases were replaced by an arbitrary orthonormal basis set. For the exponential basis set, the necessary and sufficient conditions are

Result 1 Channel identifiability is ensured iff $N \geq L(Q+1)$, the training sequence has at least L non-zero tones, and $\boldsymbol{\mathcal{C}}^H \mathbf{D}_q \boldsymbol{\mathcal{C}} = \mathbf{0}$, $q = \pm Q, ..., \pm 1$.

Let $\mathbf{D}_{\tilde{c}} = \text{diag}\left(\tilde{\boldsymbol{c}}\right)$ with $\tilde{\boldsymbol{c}}$ being the DFT of $\boldsymbol{c}$. Then a sufficient condition is that $\mathbf{D}_{\tilde{c}}^H \mathbf{J}^q \mathbf{D}_{\tilde{c}} = \mathbf{0}$, $q \neq 0$, where $\mathbf{J}$ is the circular shift matrix operator. Thus, the training sequence must satisfy an interesting shift-orthogonality in the frequency domain.

In the following, we refer to the indices of the nonzero entries of $\tilde{\boldsymbol{c}}$ as pilot frequencies. Let $\mathcal{P}$ denote the subset of $\{0, \cdots, N-1\}$ containing these pilot frequencies and let P denote its cardinality. Note that a necessary but not sufficient condition for channel identifiability is $P \geq L$.

Corollary 1 If $P = L$, then channel identifiability is guaranteed if the pilot frequencies are spaced at least $(Q+1)$ apart.

We make the following remarks

- Corollary 1 implies that $L(Q+1)$ unknown channel coefficients can be identified with only L pilot tones. When $Q = 0$, this is the standard result: the training sequence must have at least L tones to cope with the unknown possibly annihilating $L-1$ channel zeros. For $Q > 0$, this identifiability is possible thanks to the frequency diversity (or frequency spread) offered by the time-varying channel and enabled by the "shift-orthogonal" training sequence.
- A channel identifiability condition that is independent of $\mathcal{P}$ is $P \geq (L-1)Q + L$. This is required when the pilot frequencies are cyclicly contiguous.

3.2. Data-Independent Channel Estimation Condition

In order for $\hat{\boldsymbol{h}}$ to be independent of the data, the following condition must be satisfied

$$[\mathbf{I}_{Q+1} \otimes \boldsymbol{\mathcal{C}}]^H \mathbf{D}^H \mathbf{D} [\mathbf{I}_{Q+1} \otimes \underline{\mathcal{W}}] = \mathbf{0} \tag{8}$$

which can be equivalently expressed as

$$\boldsymbol{\mathcal{C}}^H \mathbf{D}_q \underline{\mathcal{W}} = \mathbf{0}, \qquad q = -Q, \cdots, Q \tag{9}$$

Using the same reasoning as in the previous subsection, condition (9) can be expressed in the frequency domain as

$$\sum_{m=0}^{N-1} \tilde{c}^*(m) \underline{\tilde{\boldsymbol{w}}}(< m+q >_N) e^{j2\pi m\ell/N} = 0$$

$$q = -Q, \cdots, Q. \ell = -L+1, \cdots, L-1 \tag{10}$$

where $\underline{\tilde{\boldsymbol{w}}}$ is the DFT of $\underline{\boldsymbol{w}}$.

Let $\mathcal{Z}$ be the subset of $\{0, \cdots N-1\}$ containing the indices of the DFT entries of $\underline{\tilde{\boldsymbol{w}}}$ involved in Condition (10), and let Z denote its cardinality. Note that $\mathcal{Z}$ depends on $\mathcal{P}$. More specifically, if $k \in \mathcal{P}$, then $\{< k-Q >_N, \cdots, < k+Q >_N\} \subset \mathcal{Z}$. Condition (10) imposes $(2L-1)(2Q+1)$ constraints on Z elements of $\underline{\tilde{\boldsymbol{w}}}$ indexed by $\mathcal{Z}$. Therefore, the number of effective constraints on $\underline{\tilde{\boldsymbol{w}}}$ (or $\underline{\boldsymbol{w}}$) is $\min(Z, (2L-1)(2Q+1))$. By keeping P to the minimum value for a given pilot placement scheme, it can be shown that $\min(Z, (2L-1)(2Q+1)) = Z$. In this case, Condition (10) implies that the $n \in \mathcal{Z}$-th DFT entries of $\underline{\tilde{\boldsymbol{w}}}$ must be set to zero. Hence, in what follows, P will be kept to the minimum value. We now make the following remarks.

- If the pilot frequencies are cyclicly contiguous, the minimal value of P that guarantees channel identifiability is $P = (L-1)Q + L$, as mentioned in the previous subsection. In this case, $\mathcal{Z}$ consists of only one cluster of size $Z = (L+1)Q + L$.
- In the case where L pilot frequencies are spaced at least $(Q+1)$ apart, as in Corollary 1, $\mathcal{Z}$ consists of L disjoint clusters of size $2Q+1$, and $Z = L(2Q+1)$. Since Z in this case is larger than that obtained in the case of contiguous pilot frequencies, data distortion is also greater. However, as we will see in the next section, designing the pilot frequencies to be contiguous is worst when performance of channel estimation is concerned.
- Note that $Z > P$ for time-varying channels, unlike the case of time-invariant channels (i.e., $Q = 0$) where $Z = P$ regardless of $\mathcal{P}$ [12].
- In the presence of a DC-offset, it is preferable that $\{\mathcal{P} + q, q = -Q/2, \cdots, Q/2\}$ does not include the zero frequency in order to decouple channel and DC-offset estimation. To make DC-offset estimation data-independent, the zero frequency should be added to $\mathcal{Z}$.

4. OPTIMUM TRAINING SEQUENCE DESIGN

In the case of purely frequency-selective channels, it was shown in [15] that designing $\boldsymbol{c}$ so that its DFT has only L non-zero entries which are equally spaced and have the same magnitude is optimal in terms of minimizing the mean square error (MSE) of the LLS channel estimate and minimizing data distortion. For doubly-selective channels, the design of $\boldsymbol{c}$ is not as simple because as we will see later, minimizing the MSE of $\hat{\boldsymbol{h}}$ under Condition (9) does not minimize data distortion and vice-versa.

4.1. Minimizing the MSE of Channel Estimate

Since $\boldsymbol{v}$ is AWGN, the MSE of $\hat{\boldsymbol{h}}$ is, under Condition (9), given by

$$\begin{aligned}\mathsf{mse}\left(\hat{\boldsymbol{h}}\right) &:= \mathrm{Tr}\left\{E\left\{(\hat{\boldsymbol{h}}-\boldsymbol{h})(\hat{\boldsymbol{h}}-\boldsymbol{h})^H\right\}\right\} \\ &= \sigma^2 \mathrm{Tr}\left\{\left([\mathbf{I}_{Q+1}\otimes \boldsymbol{\mathcal{C}}^H]\mathbf{D}^H\mathbf{D}[\mathbf{I}_{Q+1}\otimes \boldsymbol{\mathcal{C}}]\right)^{-1}\right\}.\end{aligned}$$

where σ^2 is the noise power. We have the following inequality

$$\mathsf{mse}\left(\hat{\boldsymbol{h}}\right) \geq \sigma^2[L(Q+1)]^{-1}\mathrm{Tr}\left\{[\mathbf{I}_{Q+1}\otimes \boldsymbol{\mathcal{C}}^H]\mathbf{D}^H\mathbf{D}[\mathbf{I}_{Q+1}\otimes \boldsymbol{\mathcal{C}}]\right\}$$

with equality iff

$$[\mathbf{I}_{Q+1}\otimes \boldsymbol{\mathcal{C}}^H]\mathbf{D}^H\mathbf{D}[\mathbf{I}_{Q+1}\otimes \boldsymbol{\mathcal{C}}] \propto \mathbf{I}$$

which is equivalent to

$$\boldsymbol{\mathcal{C}}^H\mathbf{D}_q\boldsymbol{\mathcal{C}} \propto \delta(q)\mathbf{I}, \qquad q = -Q, \cdots, Q.$$

Using the same reasoning as in the previous section, the above condition becomes

$$\sum_{m=0}^{N-1} \tilde{c}^*(m)\tilde{c}(< m+q >_N)e^{j2\pi m\ell/N} = \sigma_c^2\delta(q)\delta(\ell),$$
$$q = -Q, \cdots, Q; \ell = -L+1, \cdots, L-1. \qquad (12)$$

A simple design that satisfies the above condition is

$$\begin{aligned}|\tilde{c}(k)|^2 &= \frac{N\sigma_c^2}{P}\sum_{i=0}^{P-1}\delta(k-iM-t), \qquad k = 0, \cdots, N-1 \\ \text{and} & \\ N &= PM,\ 0 \leq t \leq M-1,\ P \geq L,\ M \geq Q+1 \quad (13)\end{aligned}$$

where P, t and M are positive integers and $\sigma_c^2 = (1/N)\sum_{n=0}^{N-1}|c(n)|^2$. This design consists of P equispaced tones, at least $Q+1$ apart. When $M \leq Q$, Condition (12) can still be satisfied but in this case, the phases of the $\tilde{c}(m)$ would have to be constrained as well. However, the training design in this case is not interesting since P should be minimized in the DDST approach, as we will see later. Note that in the presence of a DC offset, t should be larger than $Q/2$ in order to decouple channel and DC offset estimations [14].

Using eq. (13), the minimum MSE is given by

$$\mathsf{mse}\left(\hat{\boldsymbol{h}}\right)\Big|_{\min} = \frac{\sigma^2 L(Q+1)}{N\sigma_c^2}. \qquad (14)$$

It is worth noting that the minimum MSE is not a function of P, the number of non-zero entries of $\tilde{\boldsymbol{c}}$. Thus, P should be set to its minimum value, L, in order to minimize data distortion. Recall that when L pilot frequencies are spaced at least $(Q+1)$ apart, the number of zeroed entries of $\underline{\tilde{\boldsymbol{w}}}$ is $Z = L(2Q+1)$.

Note that for the optimal design in eq. (13), the channel estimate in eq. (6) reduces to

$$\hat{\boldsymbol{h}}_q = \frac{1}{\sigma_c^2}\boldsymbol{\mathcal{C}}^H\mathbf{D}_q^H\boldsymbol{y}, \quad q = -Q/2, \cdots, Q/2$$

The coefficients of $\hat{\boldsymbol{h}}_q$ can also be simply expressed as

$$\hat{h}_{q,\ell} = \frac{1}{N\sigma_c^2}\sum_{k=0}^{P-1}\tilde{c}^*(kM+t)r_{q,k}e^{j2\pi(kM+t)/N} \qquad (15)$$

where

$$r_{q,k} = \sum_{n=0}^{N-1} y(n)e^{-j2\pi n(q+kM+t)/N}$$

A constant amplitude sequence satisfying the optimality condition in Result 1 is given by the following shifted chirp sequence

$$c(n) = \sigma_c e^{j2\pi nt/N}e^{j\pi n(n+\nu)/P}$$

where $\nu = 0$ if P is even and $\nu = 1$ if ν is odd.

4.2. Minimizing Data Distortion

In order to minimize data-distortion, we choose $\underline{\boldsymbol{w}}$ which minimizes the Euclidean distance between $\boldsymbol{w}$ and $\underline{\boldsymbol{w}}$ under the constraint that the DFT entries of $\underline{\boldsymbol{w}}$ at the frequencies $\mathcal{Z}$ are identically zero. Using Parseval's theorem, this is equivalent to minimizing

$$\sum_{k\notin\mathcal{Z}} |\tilde{w}(k) - \underline{\tilde{w}}(k)|^2 + \sum_{k\in\mathcal{Z}} |\tilde{w}(k)|^2$$

over $\{\underline{\tilde{w}}(k),\ k \in \mathcal{Z}\}$. The minimum is obtained when $\underline{\tilde{w}}(k) = \tilde{w}(k)$ for all $k \notin \mathcal{Z}$. Thus,

$$\underline{\boldsymbol{w}} = (\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{w}$$

with $\boldsymbol{\Phi} = \mathbf{F}^H \mathbf{T}_{\mathcal{Z}} \mathbf{F}$ where $\mathbf{T}_{\mathcal{Z}}$ is obtained after setting the $k \in \mathcal{Z}$th diagonal entries of the $(N \times N)$ identity matrix to zero. The power of data distortion,

$$E\{\|\boldsymbol{w} - \underline{\boldsymbol{w}}\|_2\} = E\{\|\boldsymbol{\Phi}\boldsymbol{w}\|_2\}$$

is, under the assumption of i.i.d. data symbols, given by $Z\sigma_w^2$. Thus, data distortion increases with Z but is not a function of the placements of the zeroed DFT entries of $\boldsymbol{w}$. This implies that minimizing Z also minimizes data distortion. In Subsection 3.2., it was shown that Z is minimum when the pilot frequencies are cyclicly contiguous; $Z = (L+1)Q + L$. However, this pilot placement is not optimal for channel estimation. Recall that for the optimal pilot design in eq. (13) where $P = L$, we have that $Z = L(2Q+1)$, i.e., $(L-1)Q$ more zeroed DFT entries than the minimum value obtained with cyclicly contiguous pilot frequencies. Note that in the case of purely time-selective channels (i.e., $L = 1$), the optimum design in eq. (13) with $P = L = 1$ also minimizes Z. Indeed, in this case, the optimal $\tilde{\boldsymbol{c}}$ contains only one non-zero element at an arbitrary frequency i, $\mathcal{Z} = \{< i-Q >_N, \cdots, < i+Q >_N\}$ and $Z = 2Q + 1$.

5. LINEAR EQUALIZATION AND DATA DETECTION

Since the $\mathcal{H}_q$'s are circular matrices, they can be diagonalised using the DFT matrix, i.e. $\mathcal{H}_q = \mathbf{F}^H \mathbf{H}_q \mathbf{F}$ where $\mathbf{H}_q = \text{diag}(H_q(n),\ n = 0, \cdots, N-1)$ with $H_q(n) = \sum_{\ell=0}^{L-1} h_{q,\ell} \exp(-j2\pi\ell n/N)$. Therefore, left-multiplying $\boldsymbol{y}$ in eq. (3) by $\mathbf{F}$ and using the matrix manipulations in subsection 3.1., we obtain

$$\mathbf{F}\boldsymbol{y} = \left(\sum_{q=-Q/2}^{Q/2} \mathbf{J}^q \mathbf{H}_q\right) \mathbf{F}\boldsymbol{s} =: \mathbf{H}\mathbf{F}\boldsymbol{s}. \quad (16)$$

Thus, the MMSE equalizer of $\boldsymbol{s}$ is given by

$$\hat{\boldsymbol{s}} = \mathbf{F}^H \mathbf{G} \mathbf{F} \boldsymbol{y} \quad (17)$$

where $\mathbf{G} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \sigma_w^2 \mathbf{I})^{-1}$. The soft decision of $\underline{\boldsymbol{w}}$ is then given by

$$\hat{\underline{\boldsymbol{w}}} = \hat{\boldsymbol{s}} - \boldsymbol{c}$$

The above block MMSE equalizer can be replaced by the low-complexity approximation in [16]. Further, iterative methods such as those proposed in [17] can also be implemented. Such methods were shown to outperform MMSE equalization because they better take advantage of the frequency and time diversity of the time-varying channel.

Due to data distortion at the transmission, $\hat{\underline{\boldsymbol{w}}}$ is different from $\boldsymbol{w}$ even in the absence of channel estimation error and noise. Indeed, in this ideal scenario, $\hat{\underline{\boldsymbol{w}}} = (\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{w}$. Since $(\mathbf{I} - \boldsymbol{\Phi})$ is singular, $\boldsymbol{w}$ cannot be recovered linearly. However, using the fact that the data symbols are drawn from a finite alphabet and that $\boldsymbol{\Phi}\boldsymbol{w}$ is small compared to $\boldsymbol{w}$, symbol detection can be undertaken by finding the vector of constellation points $\boldsymbol{w}$ that minimizes the Euclidian distance between $\hat{\underline{\boldsymbol{w}}}$ and $(\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{w}$. This sequence detection scheme is computationally cumbersome. Further, if sequence detection were to be used, then maximum likelihood detection (such as sphere decoding) should be preferred to linear equalization. Here, we use the iterative symbol-by-symbol detection scheme proposed in [12].

The symbol-by-symbol detection algorithm is initialized by treating $\boldsymbol{\Phi}\boldsymbol{w}$ as an extra additive noise, and considering $\hat{\underline{\boldsymbol{w}}}$ as a soft detector of $\boldsymbol{w}$; the initial hard detector of $\boldsymbol{w}$ is given by

$$\bar{\boldsymbol{w}}^{(0)} = \lfloor \hat{\underline{\boldsymbol{w}}} \rfloor$$

where $\lfloor \hat{\underline{\boldsymbol{w}}} \rfloor$ denotes the vector of constellation points that are the closest to the vector $\hat{\underline{\boldsymbol{w}}}$. The detected symbols are used to estimate $\boldsymbol{\Phi}\boldsymbol{w}$ to be used in the next iteration. The detected symbols at the ith iteration are given by

$$\bar{\boldsymbol{w}}^{(i)} = \lfloor \boldsymbol{u} + \boldsymbol{\Phi}\bar{\boldsymbol{w}}^{(i-1)} \rfloor.$$

6. SIMULATION RESULTS

We compare the proposed DDST scheme with the TDM scheme proposed in [6] in terms of channel estimation performance and bit error rate (BER). The length of the data block is set to $N = 256$. The time-varying channel is assumed to be of order $L = 3$ and generated using Jakes model with a normalized Doppler frequency f_D. Two values of f_D are considered here: $f_D = 0.003$ and $f_D = 0.005$. The channel coefficients are assumed uncorrelated and their powers are given by the exponential delay profile $E\{|h(n;\ell)|^2\} = \exp(-0.2\ell)$, $\forall n$. The exponential basis function model for the channel is used at the receiver for channel estimation with $Q = 2\lceil f_D N \rceil$. For the values of f_D mentioned above, we have that $Q = 2$ and $Q = 4$. For both schemes, we use MMSE equalization. The training sequence for the DDST method is the shifted chirp sequence given in Section 4.1 and its power is set to 10% of the total transmit power. For the TDM method, zero-guarded pilots are uniformly placed within the block as in [6].

The merits of the two methods are assessed using 500 Monte-Carlo runs. Figure 1 show the normalized mse on channel estimation which is defined as

$$\frac{\sum_{n=0}^{N-1}\sum_{\ell=0}^{L-1} |h(n;\ell) - \hat{h}(n;\ell)|^2}{\sum_{n=0}^{N-1}\sum_{\ell=0}^{L-1} |h(n;\ell)|^2}$$

for different values of the data rate loss of the TDM method. Note that $\hat{h}(n;\ell)$ is obtained using the BEM and the $\hat{h}_{q,\ell}$'s. Figure 2 shows the BER performance. The MSE and the BER level off at high SNR because of the channel modelling mismatch due to the BEM approximation. It is seen that the proposed method outperforms the TDM method in terms of channel estimation. It also compares favorably with the TDM method in terms of the BER. Recall that the proposed method does not incur any data rate loss apart from the periodic cyclic prefix insertion.

Simulation results also show that unlike the case of time-invariant channels, the iterative scheme in the previous section does not seem to provide any significant improvement. This is due to the fact that the bit error rate at high SNR is dominated by the channel modelling mismatch.

7. CONCLUSIONS

We extended the data-dependent superimposed training scheme in [12] to time-varying channels. We have derived conditions for channel identifiability and zero-interference between pilots and data. The latter was achieved without trading off data-rate. The only penalties were a slight decrease in data-to-noise power ratio and a slight data distortion. The proposed method was shown to compare favorably with time-division multiplexing.

REFERENCES

[1] A. M. Sayeed and B. Aazhang, "Joint multipath-doppler diversity in mobile wireless communications," *IEEE Trans. Communications*, vol. 47, January 1999.

[2] X. Ma and G. B. Giannakis, "Maximum-Diversity Transmissions over Doubly-Selective Wireless Channels," *IEEE Trans. Information Theory*, vol. 49(7), July 2003.

[3] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, November 1991.

[4] A. Adireddy, L. Tong and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Information Theory*, vol. 48(8), August 2002.

[5] G. B. Giannakis and C. Tepedelenlioğlu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, November 1998.

[6] X. Ma, G.B. Giannakis and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels", *IEEE Trans. Sig. Proc.*, 51 (5), May 2003.

[7] A.P. Kannu and P. Schniter, "MSE-Optimal training for linear time-varying channels", *Proc. IEEE ICASSP 2005*, Philadelphi, USA, March 2005.

[8] M. Dong, L. Tong, and B.M. Sadler, "Optimal insertion of pilot symbols for transmission over time-varying flat fading channels", *IEEE Trans. Sig. Proc.*, 52(5), May 2004.

[9] G. Tong Zhou and Ning Chen "Superimposed training for doubly selective channels", *IEEE Stat. Sig. Proc. Workshop* 2003.

[10] J. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics", *IEEE Communications Lett.*, vol. 7(9), September 2003.

[11] A. G. Orozco-Lugo, M. M. Lara and D. C. McLernon, "Channel estimation using implicit training," *IEEE Trans. Signal Processing*, vol. 52(1), January 2004.

[12] M. Ghogho, D. C. McLernon, E. Alameda-Hernande, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training", *IEEE Signal Processing Lett.*, vol. 12(3), March 2005.

[13] M. Ghogho, D.C. McLernon, E. Alameda-Hernandez and A. Swami, "SISO and MIMO Channel Estimation and Symbol Detection using Data-Dependent Superimposed Training," *IEEE Int. Conference on Acoustics, Speech and Signal Processing* (ICASSP'05) 2005, Philadelphia, USA, March 2005.

[14] M. Ghogho, "Channel and DC-offset estimation using data-dependent superimposed training," *IEE Electronics Letters*, 41 (22), October 2005.

[15] M. Ghogho and A. Swami, "Optimal Training for Affine-Precoded and Cyclic-Prefixed Block Transmissions," *IEEE Workshop on Statistical Signal Processing* (SSP), Bordeaux, France, July 2005.

[16] L. Rugini, P. Banelli, and G. Leus, "Simple Equalization of Time-Varying Channels for OFDM," *IEEE Communications Letters*, 9(7), July 2005.

[17] P. Schniter and L. Hong, "Iterative equalization for single-carrier cyclic-prefix in doubly-dispersive channels," *Thirty-Seventh Asilomar Conference*, USA, November 2003.

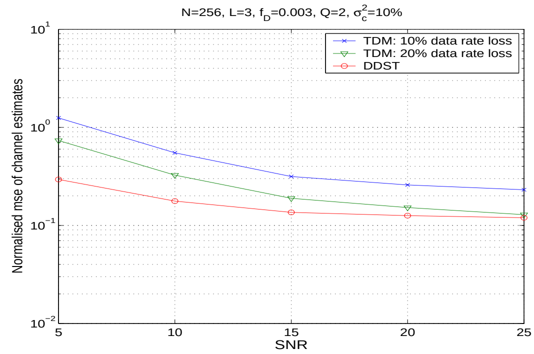

Figure 1. Empirical MSE of channel estimates

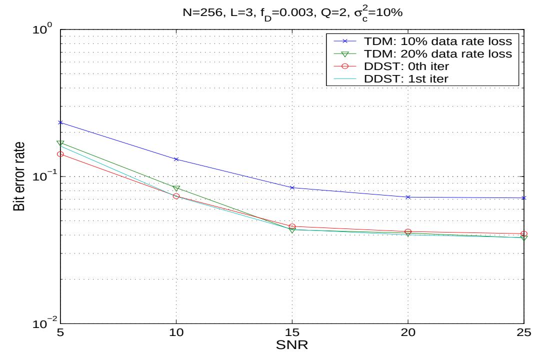

Figure 2. Empirical bit error rate

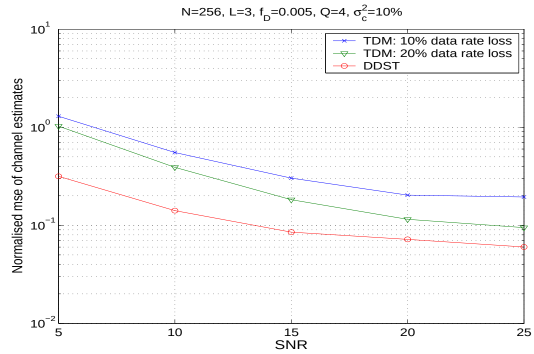

Figure 3. Empirical MSE of channel estimates

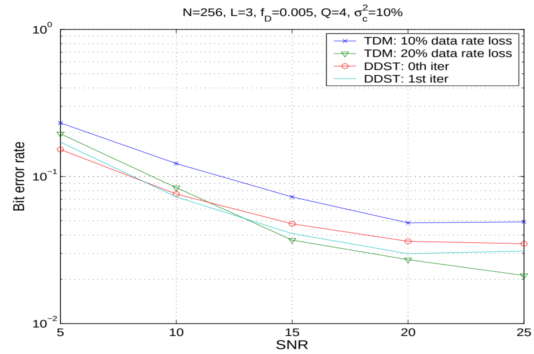

Figure 4. Empirical bit error rate