ON SUPERIMPOSED-TRAINING POWER ALLOCATION FOR TIME-VARYING CHANNEL ESTIMATION

Jitendra K. Tugnait *Shuangchi He* *Xiaohong Meng*

Department of Electrical & Computer Engineering
Auburn University, Auburn, Alabama 36849, USA
tugnajk@eng.auburn.edu heshuan@auburn.edu mengxia@auburn.edu

ABSTRACT

Channel estimation for single-input single-output (SISO) frequency-selective time-varying channels is considered using superimposed training. The time-varying channel is assumed to be described by an orthogonal polynomial basis expansion model (OP-BEM). A periodic (non-random) training sequence is arithmetically added (superimposed) at a low power to the information sequence at the transmitter before modulation and transmission. First we present a novel approach to channel estimation using only the first-order statistics of the data under a fixed power allocation to training. We then present a performance analysis of this approach for time-varying random channels to obtain a closed-form expression for the channel estimation variance. Finally, we address the issue of superimposed training power allocation. Illustrative computer simulation examples are presented where a frequency-selective channel is randomly generated with different Doppler spreads via Jakes' model.

1. INTRODUCTION

Accurate knowledge of the channel state information (CSI) of wireless communication systems is a prerequisite for most physical layer approaches. In (conventional) training-based approaches, training sequences (known to the receiver) are time-multiplexed with the information sequence and transmitted. This incurs a loss in spectral efficiency. At the receiver, one estimates the channel via least-squares and related approaches. For time-varying channels, one has to send training signals frequently and periodically to keep up with the changing channel. This wastes resources. An alternative is to estimate the channel based solely on noisy data exploiting statistical and other properties of the information sequences. This is the blind channel estimation approach. In semi-blind approaches, there is a training sequence but one uses the non-training based data also to improve the training-based results: it uses a combination of training and blind cost functions. More recently a superimposed training based approach has been explored where a training sequence is 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.

Let $\{s(n)\}$ denote a scalar sequence which is input to a SISO time-varying channel with discrete-time impulse response $\{h(n;l)\}$ (channel response at time n to a unit input at time $n-l$). Then the symbol-rate, channel output is given by

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

The noisy measurements of $x(n)$ are given by

$$y(n) = x(n) + v(n) \tag{2}$$

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

In a polynomial basis expansion representation [1] it is assumed that (following a Taylor series expansion of the doubly-selective channel around $\overline{n}$ over a fixed time interval)

$$h(n;l) = \sum_{k=1}^{K+1} \tilde{h}_k(l)(n-\overline{n})^{k-1} \tag{3}$$

where $\tilde{h}_k(l)$ (for $k = 1, 2, \cdots, K+1$) are time-invariant. Eqn. (3) is the polynomial basis expansion model (P-BEM). In OP-BEM we modify (3) to

$$h(n;l) = \sum_{q=0}^{K} h_q(l)\phi_q(n) \tag{4}$$

where $\{\phi_q(n)\}$ are an orthonormal set of basis functions (discretized modified Legendre polynomials) over the time interval of interest; see [6] for details. In a superimposed training based approach one takes

$$s(n) = b(n) + c(n), \tag{5}$$

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. Superimposed training-based approaches have been discussed in [3], [4] and [5] for SISO systems. Periodic superimposed training for channel estimation via first-order statistics for SISO systems have been discussed in [7],[9] and [11] (and reference therein) for time-invariant channels, and in [8] for both time-invariant and time-varying (CE-BEM (complex-exponential basis expansion model) based) channels.

Objectives and Contributions:

- First we present a novel approach to channel estimation using only the first-order statistics of the data under a fixed power allocation to training. Here $\{\phi_q(n)\}$'s are known and we estimate the time-invariant parameters $h_q(l)$'s for different q's and l's. We extend the approach of [2] for time-invariant channels to time-varying channels.
- We then present a performance analysis of the above proposed approach for the specific case of time-varying random channels to obtain a closed-form expression for the channel estimation variance.
- Finally, we address the issue of superimposed training power allocation. Using the developed channel estimation variance expression, we cast the power allocation problem as one of optimizing a signal-to-noise ratio (SNR) for equalizer design.

Notation: Superscripts $*$, H, T and $\dagger$ denote the complex conjugate, complex conjugate transpose, the transpose and the Moore-Penrose pseudo-inverse 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. $\mathrm{tr}(A)$ denotes the trace of matrix A.

2. PROPOSED APPROACH

In this section the channel $h(n;l)$ is assumed to be nonrandom. In Secs. 3 and 4 we will assume it to be random.

2.1. Model Assumptions

The time-varying channel $\{h(n;l)\}$ satisfies (4) where the functions $\phi_q(n)$ are known. The information sequence $\{b(n)\}$ is zero-mean, i.i.d. with $E\{|b(n)|^2\} = \sigma_b^2$. The measurement noise $\{v(n)\}$ is possibly **nonzero-mean** ($E\{v(n)\} = m$), white, uncorrelated with $\{b(n)\}$, with $E\{[v(n+\tau)-m][v(n)-m]^*\} = \sigma_v^2\delta(\tau)$. The mean m may be unknown. The superimposed training sequence $c(n) = c(n+mP)$ $\forall m, n$ is a non-random periodic sequence with period P, with average power $\sigma_c^2 := P^{-1}\sum_{n=0}^{P-1}|c(n)|^2$. Assume that the superimposed training sequence is such that $\mathcal{C}$ defined later in (12) has full column rank.

2.2. Channel Estimation

Define the "error" $e(n)$ and cost function J, respectively,

$$e(n) = y(n) - \sum_{l=0}^{L}\sum_{q=0}^{K} h_q(l)\varphi_q(n)c(n-l) - m, \quad J = \sum_{n=1}^{T}|e(n)|^2. \tag{6}$$

Choose m and $h_q(l)$'s ($q = 0, 1, \cdots, K;\ l = 0, 1, \cdots, L$) to minimize J. For optimization, we must have

$$\left.\frac{\partial J}{\partial m^*}\right|_{\substack{m=\hat{m}\\ h_q(l)=\hat{h}_q(l)\\ \forall q,l}} = 0\,, \quad \left.\frac{\partial J}{\partial h_q^*(l)}\right|_{\substack{m=\hat{m}\\ h_q(l)=\hat{h}_q(l)\\ \forall q,l}} = 0\,. \tag{7}$$

The above equations lead to

$$\hat{m} = \frac{1}{T}\sum_{n=1}^{T}y(n) - \frac{1}{T}\sum_{n=1}^{T}\sum_{l=0}^{L}\sum_{q=0}^{K}\hat{h}_q(l)\varphi_q(n)c(n-l) \tag{8}$$

and, for $0 \le q_1 \le K$, $0 \le l_1 \le L$,

$$\sum_{l=0}^{L}\sum_{q=0}^{K}\hat{h}_q(l)\left[\frac{1}{T}\sum_{n=1}^{T}\varphi_q(n)\varphi_{q_1}^*(n)c(n-l)c^*(n-l_1)\right]$$

$$= \frac{1}{T}\sum_{n=1}^{T}y(n)\varphi_{q_1}^*(n)c^*(n-l_1) - \frac{\hat{m}}{T}\sum_{n=1}^{T}\varphi_{q_1}^*(n)c^*(n-l_1). \tag{9}$$

Substitute (8) in (9), and we have

$$\sum_{l=0}^{L}\sum_{q=0}^{K}\hat{h}_q\left(l\right)\phi\left(q, q_1, l, l_1\right)$$

$$= \frac{1}{T}\sum_{n=1}^{T}y(n)\left[\varphi_{q_1}^*(n)c^*\left(n-l_1\right) - \frac{1}{T}\sum_{n=1}^{T}\varphi_{q_1}(n)c\left(n-l_1\right)\right] \tag{10}$$

where $0 \le q_1 \le K$, $0 \le l_1 \le L$,

$$\phi\left(q, q_1, l, l_1\right) := \frac{1}{T}\sum_{n=1}^{T}\varphi_q(n)\varphi_{q_1}^*(n)c\left(n-l\right)c^*\left(n-l_1\right)$$

$$-\left[\frac{1}{T}\sum_{n=1}^{T}\varphi_q(n)c\left(n-l\right)\right]\left[\frac{1}{T}\sum_{n=1}^{T}\varphi_{q_1}^*(n)c^*\left(n-l_1\right)\right]. \tag{11}$$

To rewrite (10) in a matrix expression, let

$$\mathcal{C} := \begin{bmatrix}
\phi(0,0,0,0) & \phi(1,0,0,0) & \cdots & \phi(K,0,0,0) & \phi(0,0,1,0) & \cdots & \phi(K,0,1,0) & \cdots & \phi(K,0,L,0)\\
\phi(0,1,0,0) & \phi(1,1,0,0) & \cdots & \phi(K,1,0,0) & \phi(0,1,1,0) & \cdots & \phi(K,1,1,0) & \cdots & \phi(K,1,L,0)\\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\
\phi(0,K,0,0) & \phi(1,K,0,0) & \cdots & \phi(K,K,0,0) & \phi(0,K,1,0) & \cdots & \phi(K,K,1,0) & \cdots & \phi(K,K,L,0)\\
\phi(0,0,0,1) & \phi(1,0,0,1) & \cdots & \phi(K,0,0,1) & \phi(0,0,1,1) & \cdots & \phi(K,0,1,1) & \cdots & \phi(K,0,L,1)\\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\
\phi(0,K,0,1) & \phi(1,K,0,1) & \cdots & \phi(K,K,0,1) & \phi(0,K,1,1) & \cdots & \phi(K,K,1,1) & \cdots & \phi(K,K,L,1)\\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\
\phi(0,K,0,L) & \phi(1,K,0,L) & \cdots & \phi(K,K,0,L) & \phi(0,K,1,L) & \cdots & \phi(K,K,1,L) & \cdots & \phi(K,K,L,L)
\end{bmatrix} \tag{12}$$

which is a matrix of dimensions $(K+1)(L+1) \times (K+1)(L+1)$ with its $((K+1)l_1 + q_1 + 1, (K+1)l + q + 1)$-th entry as $\phi(q, q_1, l, l_1)$. Also let a $(K+1)(L+1) \times 1$ column vector

$$\mathcal{B}(n) := \begin{bmatrix}
\varphi_0^*(n)c^*(n) - \frac{1}{T}\sum_{n=1}^{T}\varphi_0^*(n)c^*(n)\\
\varphi_1^*(n)c^*(n) - \frac{1}{T}\sum_{n=1}^{T}\varphi_1^*(n)c^*(n)\\
\vdots\\
\varphi_K^*(n)c^*(n) - \frac{1}{T}\sum_{n=1}^{T}\varphi_K^*(n)c^*(n)\\
\varphi_0^*(n)c^*(n-1) - \frac{1}{T}\sum_{n=1}^{T}\varphi_0^*(n)c^*(n)\\
\vdots\\
\varphi_K^*(n)c^*(n-1) - \frac{1}{T}\sum_{n=1}^{T}\varphi_K^*(n)c^*(n)\\
\vdots\\
\varphi_K^*(n)c^*(n-L) - \frac{1}{T}\sum_{n=1}^{T}\varphi_K^*(n)c^*(n-L)
\end{bmatrix}. \tag{13}$$

Further define

$$\hat{\mathcal{H}}_l := \begin{bmatrix} \hat{h}_0(l) & \hat{h}_1(l) & \cdots & \hat{h}_K(l) \end{bmatrix}^T$$

$$\hat{\mathcal{H}} := \begin{bmatrix} \hat{\mathcal{H}}_0^T & \hat{\mathcal{H}}_1^T & \cdots & \hat{\mathcal{H}}_L^T \end{bmatrix}^T$$

as the estimator of $h_q(l)$'s for $q = 0, 1, \cdots, K$, and $l = 0, 1, \cdots, L$; define $\mathcal{H}_l$ and $\mathcal{H}$ as the corresponding vector of true values. The matrix expression for (10) is

$$\mathcal{C}\hat{\mathcal{H}} = \frac{1}{T}\sum_{n=1}^{T}\mathcal{B}(n)y(n) \tag{14}$$

which yields

$$\hat{\mathcal{H}} = \frac{1}{T}\mathcal{C}^{\dagger}\sum_{n=1}^{T}\mathcal{B}(n)y(n). \tag{15}$$

The estimate of the time-varying channel $h(n;l)$ is then given by

$$\hat{h}(n;l) = \sum_{q=0}^{K}\hat{h}_q(l)\phi_q(n). \tag{16}$$

3. PERFORMANCE ANALYSIS FOR TIME-SELECTIVE RAYLEIGH CHANNELS

Assume that the time-varying channel $\{h(n;l)\}$ is zero-mean, complex Gaussian with variance σ^2, and is mutually independent for distinct l's: $E\{|h(n;l)|^2\} = \sigma^2$, $E\{h(n_1;l_1)h^*(n_2;l_2)\} = 0$ for $l_1 \neq l_2$, $\forall n_1, n_2$. That is, different channel taps are indepenednt of each other and are identically distributed zero-mean complex Gaussian.

We define

$$\tilde{x}(n) := y(n) - E\{y(n) \mid \mathcal{H}\} = \sum_{l=0}^{L} h(n;l)\, b(n-l) + v(n) - m.$$

Then

$$E\{\tilde{x}(n_1)\tilde{x}^*(n_2) \mid \mathcal{H}\} = \sum_{l_1=0}^{L}\sum_{l_2=0}^{L} h(n_1;l_1)\, h^*(n_2;l_2)$$

$$\cdot \sigma_b^2 \delta(n_1 - n_2 - l_1 + l_2) + \sigma_v^2 \delta(n_1 - n_2). \qquad (17)$$

Since $h(n;l)$'s are independent for different l's,

$$E_{\mathcal{H}}\{E\{\tilde{x}(n_1)\tilde{x}^*(n_2) \mid \mathcal{H}\}\}$$

$$= E\left\{\sum_{l=0}^{L} h(n_1;l)\, h^*(n_2;l)\right\} \sigma_b^2 \delta(n_1 - n_2) + \sigma_v^2 \delta(n_1 - n_2)$$

$$= \left[(L+1)\sigma^2\sigma_b^2 + \sigma_v^2\right] \delta(n_1 - n_2). \qquad (18)$$

From (15) and (18), we have

$$E_{\mathcal{H}}\left\{\mathrm{cov}\left\{\hat{\mathcal{H}}, \hat{\mathcal{H}} \mid \mathcal{H}\right\}\right\}$$

$$= \frac{1}{T^2}\mathcal{C}^{\dagger}\left[\sum_{n_1=1}^{T}\sum_{n_2=1}^{T} \mathcal{B}(n_1) E_{\mathcal{H}}\{E[\tilde{x}(n_1)\tilde{x}^*(n_2) \mid \mathcal{H}]\} \mathcal{B}^H(n_2)\right]\mathcal{C}^{\dagger H}$$

$$= \frac{\left[(L+1)\sigma^2\sigma_b^2 + \sigma_v^2\right]}{T^2}\mathcal{C}^{\dagger}\left[\sum_{n=1}^{T}\mathcal{B}(n)\mathcal{B}^H(n)\right]\mathcal{C}^{\dagger H}. \qquad (19)$$

It turns out that

$$\frac{1}{T}\sum_{n=1}^{T}\mathcal{B}(n)\mathcal{B}^H(n) = \mathcal{C} \text{ and } \mathcal{C} = \mathcal{C}^H.$$

Further define the normalized training sequence $\bar{c}(n) = \sigma_c^{-1}c(n)$ with unit average power $\frac{1}{P}\sum_{p=1}^{P}|\bar{c}(n)|^2 = 1$ and $\bar{\mathcal{C}} = \sigma_c^{-2}\mathcal{C}$. Then (19) becomes

$$E_{\mathcal{H}}\left\{\mathrm{cov}\left\{\hat{\mathcal{H}}, \hat{\mathcal{H}} \mid \mathcal{H}\right\}\right\} = \frac{\left[(L+1)\sigma^2\sigma_b^2 + \sigma_v^2\right]}{T\sigma_c^2}\bar{\mathcal{C}}^{\dagger}. \qquad (20)$$

Let

$$\Phi(n) := [\ \varphi_0(n) \quad \varphi_1(n) \quad \cdots \quad \varphi_K(n)\ ],$$

$$\Psi(n) = I_{L+1} \otimes \Phi(n).$$

From orthonormality of $\{\varphi_i(n),\ 1 \le n \le T\}_{i=0}^{K}$

$$\sum_{n=1}^{T}\Psi^H(n)\Psi(n) = I_{(L+1)(K+1)}. \qquad (21)$$

By the OP-BEM (4) and (16)

$$h(n;l) = \Phi(n)\mathcal{H}_l \text{ and } \hat{h}(n;l) = \Phi(n)\hat{\mathcal{H}}_l.$$

The channel estimator variance/MSE (mean-square error) is given by

$$\sigma_{\hat{\mathbf{h}}}^2(n) := \sum_{l=0}^{L} E_{\mathcal{H}}\left\{E\left\{\left|\hat{h}(n;l) - h(n;l)\right|^2 \mid \mathcal{H}\right\}\right\}$$

$$= \mathrm{tr}\left\{\Psi(n)E_{\mathcal{H}}\left\{\mathrm{cov}\left\{\hat{\mathcal{H}}, \hat{\mathcal{H}} \mid \mathcal{H}\right\}\right\}\Psi^H(n)\right\}$$

$$= \frac{(L+1)\sigma^2\sigma_b^2 + \sigma_v^2}{T\sigma_c^2}\mathrm{tr}\left\{\Psi(n)\bar{\mathcal{C}}^{\dagger}\Psi^H(n)\right\}. \qquad (22)$$

By (21), the time-average of $\sigma_{\hat{\mathbf{h}}}^2(n)$ over length T is

$$\bar{\sigma}_{\hat{\mathbf{h}}}^2 := T^{-1}\sum_{n=1}^{T}\sigma_{\hat{\mathbf{h}}}^2(n) = \frac{(L+1)\sigma^2\sigma_b^2 + \sigma_v^2}{T^2\sigma_c^2}\mathrm{tr}\bar{\mathcal{C}}^{\dagger}. \qquad (23)$$

4. TRAINING POWER ALLOCATION

Here we follow the time-invariant results of [10]. Define the training power overhead β as

$$\beta := \frac{\frac{1}{P}\sum_{n=1}^{P}|c(n)|^2}{\frac{1}{P}\sum_{n=1}^{P}E\left\{|s(n)|^2\right\}} = \frac{\sigma_c^2}{\sigma_b^2 + \sigma_c^2}. \qquad (24)$$

For a fixed SNR or transmitted power budget, higher β implies smaller effective SNR at the receiver due to decreased power in the information sequence but higher channel estimation accuracy. Removing the estimated time-varying mean from the received data, define

$$\tilde{y}(n) := y(n) - \sum_{l=0}^{L}\hat{h}(n;l)\, c(n-l) - \hat{m}$$

where $\hat{m}$ is given by (8). Assuming that $\hat{m} = m$, we have

$$\tilde{y}(n) \approx \underbrace{\sum_{l=0}^{L}\hat{h}(n;l)\, b(n-l)}_{=:x_s(n)}$$

$$+ \underbrace{\sum_{l=0}^{L}\left[h(n;l) - \hat{h}(n;l)\right]\left[b(n-l) + c(n-l)\right] + \tilde{v}(n)}_{=:w(n)}. \qquad (25)$$

When using $\hat{h}(n;l)$ for equalization/detection, effective noise (as a first-order approximation) is $w(n)$ whose variance contains channel estimation error variance as a component, which in turn will depend on β, and effective signal is $x_s(n)$. An "optimum" value of the β for the superimposed training method may be obtained by maximizing the SNR in (25) w.r.t. β, where this SNR is defined as

$$\mathrm{SNR}_e(\beta) = \sigma_{xs}^2/\sigma_w^2 \qquad (26)$$

under the constraint of a fixed transmitted power, leading to the constraint $\sigma_b^2 + \sigma_c^2 = P_T$. In (25), the signal power is given by

$$\sigma_{xs}^2(n) = E\left\{|x_s(n)|^2\right\} = \sigma_b^2\sum_{l=0}^{L}E_{\mathcal{H}}\left\{E\left\{\left|\hat{h}(n;l)\right|^2 \mid \mathcal{H}\right\}\right\}$$

$$= \sigma_b^2\sum_{l=0}^{L}\left[E_{\mathcal{H}}\left\{E\left\{\left|\hat{h}(n;l) - h(n;l)\right|^2 \mid \mathcal{H}\right\}\right\}\right.$$

$$+E\left\{|h(n;l)|^2\right\}\Big]$$

$$=\sigma_b^2\left[\sigma_{\hat{\mathbf{h}}}^2(n)+(L+1)\,\sigma^2\right] \quad (27)$$

and the noise power is given by

$$\sigma_w^2(n)=E\left\{|w(n)|^2\right\}$$

$$=\sigma_b^2\sigma_{\hat{\mathbf{h}}}^2(n)+\sigma_v^2+\sigma_c^2\sum_{l_1=0}^{L}\sum_{l_2=0}^{L}E\left\{\left[\hat{h}(n;l_1)-h(n;l_1)\right]\right.$$

$$\left.\cdot\left[\hat{h}(n;l_2)-h(n;l_2)\right]^*\right\}\bar{c}(n-l_1)\,\bar{c}^*(n-l_2). \quad (28)$$

Eq. (28) can be further simplified as

$$\sigma_w^2(n)=\sigma_b^2\sigma_{\hat{\mathbf{h}}}^2(n)+\sigma_v^2+\frac{\left[(L+1)\,\sigma^2\sigma_b^2+\sigma_v^2\right]}{T}$$

$$\cdot\mathrm{tr}\left\{\bar{\mathcal{C}}^\dagger\left[\mathcal{E}^H(n)\Phi^H(n)\Phi(n)\mathcal{E}(n)\right]\right\},$$

where

$$\mathcal{E}(n):=\left[\begin{array}{cccc}\bar{c}(n) & \bar{c}(n-1) & \cdots & \bar{c}(n-L)\end{array}\right]\otimes I_{K+1}.$$

Taking the time average of the signal and noise powers, we have

$$\bar{\sigma}_{xs}^2:=\frac{1}{T}\sum_{n=1}^{T}\sigma_{xs}^2(n)=\sigma_b^2\left[\bar{\sigma}_{\hat{\mathbf{h}}}^2+(L+1)\,\sigma^2\right] \quad (29)$$

$$\bar{\sigma}_w^2:=\frac{1}{T}\sum_{n=1}^{T}\sigma_w^2(n)=\sigma_b^2\bar{\sigma}_{\hat{\mathbf{h}}}^2+\sigma_v^2+\frac{\left[(L+1)\,\sigma^2\sigma_b^2+\sigma_v^2\right]}{T^2}$$

$$\cdot\mathrm{tr}\left\{\bar{\mathcal{C}}^\dagger\sum_{n=1}^{T}\left[\mathcal{E}^H(n)\Phi^H(n)\,\Phi(n)\mathcal{E}(n)\right]\right\}$$

$$=\sigma_b^2\bar{\sigma}_{\hat{\mathbf{h}}}^2+\sigma_v^2+\frac{\left[(L+1)\,\sigma^2\sigma_b^2+\sigma_v^2\right]}{T^2}\mathrm{tr}\left\{\bar{\mathcal{C}}^\dagger\mathcal{D}\right\} \quad (30)$$

where

$$\mathcal{D}:=\sum_{n=1}^{T}\mathcal{E}^H(n)\Phi^H(n)\,\Phi(n)\mathcal{E}(n)$$

and its $((K+1)\,l_1+q_1+1,(K+1)\,l+q+1)$-th entry is $\sum_{n=1}^{T}\varphi_q(n)\varphi_{q_1}^*(n)\bar{c}(n-l)\,\bar{c}^*(n-l_1)$. Furthermore, we define the time average version of (26) as

$$\mathrm{SNR}_d(\beta)=\frac{\bar{\sigma}_{xs}^2}{\bar{\sigma}_w^2}. \quad (31)$$

With β as in (24), using the constraint $\sigma_b^2+\sigma_c^2=P_T$, we have $\sigma_c^2=\beta P_T$ and $\sigma_b^2=(1-\beta)P_T$. Thus, incorporating these constraint-carrying variables in (31) via (30), (29) and (23), we have (after some manipulations) an unconstrained cost

$$\mathrm{SNR}_d(\beta)=\frac{f_1\beta^2+f_2\beta+f_3}{g_1\beta^2+g_2\beta+g_3} \quad (32)$$

where

$$f_1=(L+1)\,\sigma^2\left(\mathrm{tr}\bar{\mathcal{C}}^\dagger-T^2\right),$$

$$f_2=-(L+1)\,\sigma^2\left(2\mathrm{tr}\bar{\mathcal{C}}^\dagger-T^2\right)-\frac{\sigma_v^2\mathrm{tr}\bar{\mathcal{C}}^\dagger}{P_T},$$

$$f_3=(L+1)\,\sigma^2\mathrm{tr}\bar{\mathcal{C}}^\dagger+\frac{\sigma_v^2\mathrm{tr}\bar{\mathcal{C}}^\dagger}{P_T},$$

$$g_1=(L+1)\,\sigma^2\left(\mathrm{tr}\bar{\mathcal{C}}^\dagger-\mathrm{tr}\left\{\bar{\mathcal{C}}^\dagger\mathcal{D}\right\}\right),$$

$$g_2=-(L+1)\sigma^2\mathrm{tr}\left\{2\bar{\mathcal{C}}^\dagger-\bar{\mathcal{C}}^\dagger\mathcal{D}\right\}+\frac{\mathrm{tr}\left\{\bar{\mathcal{C}}^\dagger\mathcal{D}-\bar{\mathcal{C}}^\dagger\right\}+T^2}{P_T/\sigma_v^2},$$

$$g_3=f_3.$$

We seek the optimum value of β by taking the derivative

$$\frac{d\left[\mathrm{SNR}_d(\beta)\right]}{d\beta}$$

$$=\frac{(f_1g_2-f_2g_1)\,\beta^2+2\,(f_1g_3-f_3g_1)\,\beta+f_2g_3-f_3g_2}{\left(g_1\beta^2+g_2\beta+g_3\right)^2}=0,$$

the root of which lying in $[0,1]$ is

$$\beta_{\mathrm{opt}}=\frac{1}{f_1g_2-f_2g_1}\left(-f_1g_3+f_3g_1-\sqrt{a}\right) \quad (33)$$

where

$$a:=-f_1f_2g_2g_3-2f_1f_3g_1g_3-f_2f_3g_1g_2+f_2^2g_1g_3$$

$$+f_1f_3g_2^2+f_1^2g_3^2+f_3^2g_1^2.$$

4.1. Some Approximations

Since $h(n;l)$'s are mutually independent for different l's, the calculation can be further simplified if we suppose $\hat{h}(n;l)$'s are also approximately uncorrelated for distinct l's, i.e.,

$$E\left\{\left[\hat{h}(n;l_1)-h(n;l_1)\right]\left[\hat{h}(n;l_2)-h(n;l_2)\right]^*\right\}\approx 0,\ l_1\neq l_2.$$

Then (28) becomes

$$\sigma_w^2(n)\approx\sigma_b^2\sigma_{\hat{\mathbf{h}}}^2(n)+\sigma_v^2$$

$$+\sigma_c^2\sum_{l=0}^{L}E\left\{\left|\hat{h}(n;l)-h(n;l)\right|^2\right\}\bar{c}^*(n-l)\,\bar{c}(n-l). \quad (34)$$

In addition, if $|\bar{c}(n)|^2=$ constant $\forall n$ (e.g., $\bar{c}(n)=\pm 1$), (34) can be further simplified as

$$\sigma_w^2(n)\approx\sigma_{\hat{\mathbf{h}}}^2(n)\left(\sigma_b^2+\sigma_c^2\right)+\sigma_v^2.$$

Then

$$\bar{\sigma}_{xs}^2=\sigma_b^2\left[\bar{\sigma}_{\hat{\mathbf{h}}}^2+(L+1)\,\sigma^2\right],\quad \bar{\sigma}_w^2=\bar{\sigma}_{\hat{\mathbf{h}}}^2\left(\sigma_b^2+\sigma_c^2\right)+\sigma_v^2,$$

and the unconstrained cost is

$$\mathrm{SNR}_d(\beta)=\frac{\bar{\sigma}_{xs}^2}{\bar{\sigma}_w^2}=\frac{f_1\beta^2+f_2\beta+f_3}{g_1'\beta+g_2'} \quad (35)$$

where f_1, f_2 and f_3 are as in (32), and

$$g_1'=\frac{\sigma_v^2T^2}{P_T}-(L+1)\,\sigma^2\mathrm{tr}\bar{\mathcal{C}}^\dagger,$$

$$g_2'=(L+1)\,\sigma^2\mathrm{tr}\bar{\mathcal{C}}^\dagger+\frac{\sigma_v^2\mathrm{tr}\bar{\mathcal{C}}^\dagger}{P_T}.$$

Setting the first derivative of $\mathrm{SNR}_d(\beta)$ to be zero to get the optimum β, we have

$$\beta_{\mathrm{opt}}=\frac{g_2'}{g_1'}\left[-1+\sqrt{1+\frac{g_1'\,(f_3g_1'-f_2g_2')}{g_2'^2f_1}}\right]. \quad (36)$$

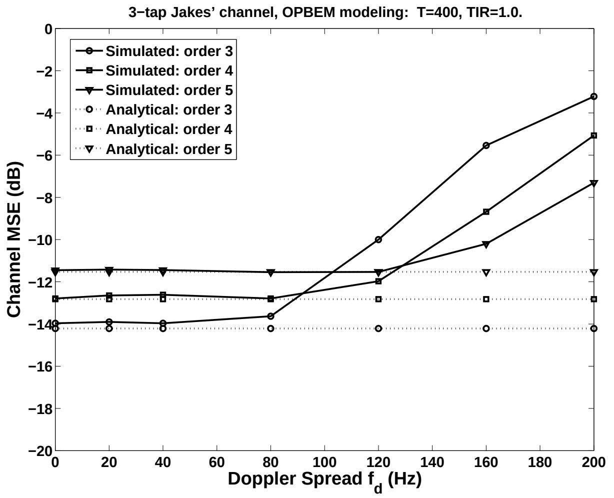

Figure 1. Normalized channel estimation variance/MSE (in dB) vs Doppler spread f_d Hz for OP-BEM representation. Dashed line: theoretical (analytical) expression; solid line: simulation result. Record length = 400 bits. Results based on 100 Monte Carlo runs. (Polynomial orders refer to the value of K in OP-BEM.)

5. SIMULATION EXAMPLES

We simulate a 3-tap ($L = 2$) channel with each tap following a Jakes' model (see [13]) with a specified Doppler spread f_d. Note that Jakes' model doesn't necessarily follow (4), i.e., there are model approximation errors. We consider a system with carrier frequency of 2GHz, data rate of 40kB (kB=kilo-Bauds), therefore, $T_s = 25 \times 10^{-6}$ sec., and a varying Doppler spread f_d in the range 0 to 200Hz (corresponding to a maximum mobile velocity in the range 0 to 108km/hr). We set $E\left\{|h(n;l)|^2\right\} = \sigma^2 = 1/(L+1)$. We take all sequences (information and training) to be binary. For superimposed training, we take a periodic (scaled) binary m-sequence of period $P = 7$ (it $(\{\bar{c}(n)\})$ is a scaled m-sequence $\{1, -1, -1, 1, 1, 1, -1\}$) with the training-to-information sequence power ratio (TIR) of 1.0 where TIR= σ_c^2/σ_b^2. Fig. 1 shows the normalized channel mean-square error (NCMSE) in channel estimation as results of our proposed approach, defined as

$$\mathrm{NCMSE} = \frac{T^{-1}\sum_{n=0}^{T-1}\sum_{l=0}^{2}\left|\hat{h}(n;l) - h(n;l)\right|^2}{E\left\{|h(n;l)|^2\right\}}$$

obtained via simulations by averaging over 100 Monte Carlo runs (different channel in each run) and via (time-averaged) analytical expression (23). An excellent agreement is seen at lower Doppler spreads whereas model approximation errors become significant at higher Doppler spreads.

In Fig. 2 and Fig. 3, a Viterbi detector is applied based on the estimated channel to detect the information sequence, and its performance is studied with the estimated channel responses. Fig. 2 shows the performance for different received SNR's with a fixed Doppler spread of 20Hz. The optimum value of β increases with increasing SNR.

Fig. 3 compares the performance of the Viterbi detector with different Doppler spreads f_d. The results of Fig. 3 show the efficacy of the proposed approach in dealing with different time-varying channels. For Doppler spreads lower

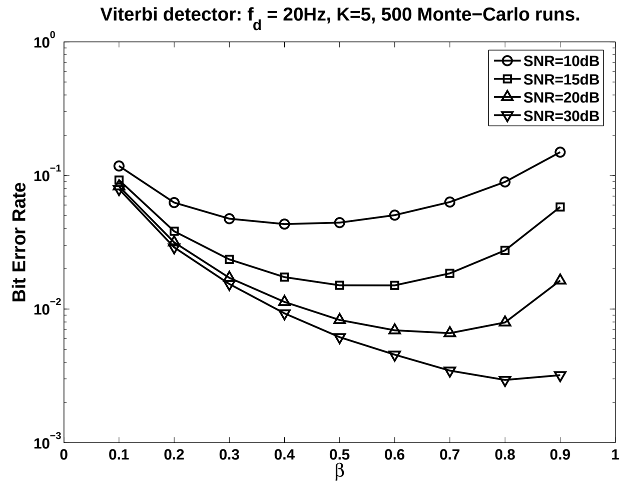

Figure 2. Bit error rate of Viterbi detector vs β with $K = 5$, $f_d = 20$Hz, $T = 400$.

than 160Hz, the BER does not change much implying that the channel estimation based on OP-BEM works well.

In Fig. 4 we compare the optimum values of β for a given received signal SNR for three different values of the Doppler spread and for three cases: that maximizing theoretical $\mathrm{SNR}_d(\beta)$ (32) (labeled "theoretical" in Fig. 4), that maximizing the approximate theoretical $\mathrm{SNR}_d(\beta)$ (35) (labeled "approximate theoretical" in Fig. 4), and that maximizing the BER based on the Viterbi detector (labeled "Viterbi det." in Fig. 4). It is seen that all the curves have the same trend, i.e., optimum β increases as SNR increases. The approximate theoretical expression provides good solutions for β for practical applications.

6. CONCLUSIONS

We presented a novel approach to channel estimation using only the first-order statistics of the data under a fixed power allocation to training. We extended the approach of [2] for time-invariant channels to time-varying channels. By using OP-BEM, our approach can be efficiently applied to a wide range of time-varying channels. We then presented a performance analysis of the above proposed approach for the specific case of time-varying random channels to obtain a closed-form expression for the channel estimation variance. Next, we addressed the issue of superimposed training power allocation. Using the developed channel estimation variance expression, we cast the power allocation problem as one of optimizing a signal-to-noise (SNR) for equalizer design. Simulation examples were presented in support of the proposed approaches and analysis.

Although the details were provided for OP-BEM, our results apply to any orthonormal basis functions, such as CE-BEM.

7. REFERENCES

[1] D. Borah and B. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model" *IEEE Tran. Commun.*, vol Com-47, pp. 862-873, June 1999.

[2] B. Frahang-Boroujeny, "Pilot-based channel identification: proposal for semi-blind identification of communications channels" *Electronics Letters*, vol. 31, No. 15, pp. 1044-1046, June 1995.

[3] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proc. IEEE GLOBECOM*

Conf., pp. 2162-66, Rio de Janeiro, Brazil, Dec. 5-9, 1999.

[4] T.P. Holden and K. Feher, "A spread spectrum based system technique for synchronization of digital mobile communication systems," *IEEE Trans. Broadcasting,* vol. 36, pp. 185-194, Sept. 1990.

[5] F. Mazzenga, "Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," *IEEE Trans. Broadcasting,* vol. 46, pp. 170-176, June 2000.

[6] Xiaohong Meng and J.K. Tugnait, "Superimposed training-based doubly-selective channel estimation using exponential and polynomial bases models," in *Proc. 2004 Conf. Inform. Sciences & Systems*, Princeton University, NJ, March 17-19, 2004.

[7] A.G. Orozco-Lugo, M.M. Lara and D.C. McLeron, "Channel estimation using implicit training," *IEEE Trans. Signal Proc.,* vol. SP-52, pp. 240-254, Jan. 2004.

[8] J.K. Tugnait and Weilin Luo, "On channel estimation using superimposed training and first-order statistics," in *Proc. 2003 ICASSP,* Hong Kong, April 2003.

[9] J.K. Tugnait and Weilin Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Letters,* vol. CL-8, pp. 413-415, Sept. 2003.

[10] J.K. Tugnait and Xiaohong Meng, "Performance analysis and training power allocation for channel estimation using superimposed training," in *Proc. 2005 IEEE ICASSP*, Philadelphia, PA, March 18-23, 2005.

[11] G.T. Zhou, M. Viberg and T. McKelvey, "A first-order statistical method for channel estimation," *IEEE Signal Proc. Letters,* vol. SPL-10, pp. 57-60, March 2003.

[12] T.K. Moon and W.C. Stirling, *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall, 2000.

[13] Y.R. Zheng and C. Xiao, "Simulation models with correct statistical proerties for Rayleigh fading channels," *IEEE Trans. Commun.*, COM-51, pp. 920-928, June 2003.

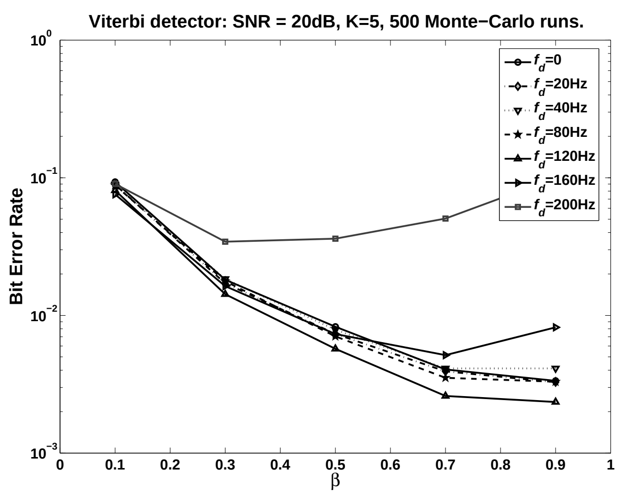

Figure 3. Bit error rate of Viterbi detector vs β with $K = 5$, SNR= 20dB, $T = 400$.

8. APPENDIX: OP-BEM REPRESENTATION

We first consider the polynomial model of [1]. Let T_s denote the symbol interval. For $t \in [t_0, t_0 + TT_s]$, expand the continuous-time baseband channel impulse response $h(t;\tau)$ in t about a midpoint $t = \overline{n}T_s + t_0$ as

$$h(t;\tau) = \sum_{i=0}^{K} \alpha_{\overline{n}}^{(i)}(\tau) \left(\frac{t - \overline{n}T_s - t_0}{T_s}\right)^i + R_K(t;\tau) \qquad (37)$$

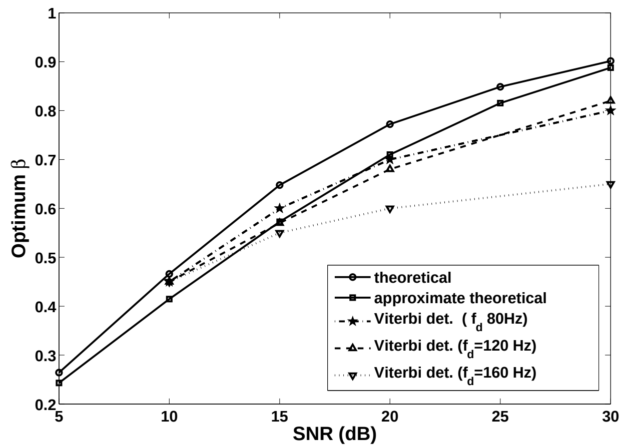

Figure 4. Optimum β vs SNR with $K = 5$, $T = 400$.

where the coefficients $\alpha_{\overline{n}}^{(i)}(\tau)$ are defined as

$$\alpha_{\overline{n}}^{(i)}(\tau) := \frac{T_s^i}{i!}\left[\frac{d^i h(t;\tau)}{dt^i}\right]_{t=\overline{n}T_s+t_0} \qquad (38)$$

and $R_K(t;\tau)$ is the remainder of the Taylor series, given by

$$R_K(t;\tau) := \frac{(t - \overline{n}T_s - t_0)^{K+1}}{(K+1)!}\left[\frac{d^{K+1}h(t;\tau)}{dt^{K+1}}\right]_{t=s'} \qquad (39)$$

for some s' lying between t and $\overline{n}T_s + t_0$. If $x(t)$ is sampled once every T_s sec. (symbol rate), then 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 given by (1) with $h(n;l)$ specified by (3) and $\overline{n} := \lceil \frac{T}{2} \rceil$. The above representation is valid over a duration of TT_s seconds (T samples).

The set of functions $\{1, t, t^2, \cdots, t^K\}$ used in (37) are linearly independent over $[-1, 1]$, but not necessarily orthogonal. The equation (37) (with $R_K(t;\tau)$ neglected) can be re-expressed in terms of orthogonal (hence orthonormal) polynomials obtained via the Gram-Schmidt procedure. Over the interval $[-1, 1]$, one gets the Legendre polynomials [12]. By appropriate scaling and translation of the (original) Legendre polynomials, we can obtain modified Legendre polynomials which are orthonormal over the interval $[t_0, t_0 + TT_s]$. Sampling these polynomials at the symbol interval, one gets the orthogonal polynomial model (OP-BEM) used in this paper. Let $p_k(\tilde{t})$ denote the orthonormal Legendre polynomial of degree (order) k over the interval $[-1, 1]$. To extend $[-1, 1]$ to $[t_0, t_0 + TT_s]$, we set $t = (TT_s/2)\tilde{t} + t_0 + (TT_s/2)$, leading to $\tilde{t} = [2/(TT_s)](t - t_0) - 1$ and modified Legendre polynomials $p'_k(t) = p_k((2/(TT_s))[t - t_0] - 1)$ orthonormal over the interval $[t_0, t_0 + TT_s]$. Sample $p'_k(t)$'s at $t = nT_s + t_0$, $(n = 0, 1, \cdots)$ to obtain the discretized modified Legendre polynomials

$$\phi_k(n) := p_k(\frac{2n}{T} - 1). \qquad (40)$$

Using the basis functions of (40), we rewrite (3) as

$$h(n;l) = \sum_{q=0}^{K} h_q(l)\phi_q(n). \qquad (41)$$