PERFORMANCE ANALYSIS AND TRAINING POWER ALLOCATION FOR CHANNEL ESTIMATION USING SUPERIMPOSED TRAINING

Jitendra K. Tugnait *Xiaohong Meng*

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

ABSTRACT

Channel estimation for single-input multiple-output (SIMO) time-invariant channels using superimposed training has been recently considered by several authors. In particular, in Tugnait and Luo (2003), the channel is estimated using only the first-order statistics of the data under a fixed power allocation to training. We first present a performance analysis of the approach of Tugnait and Luo (2003) to obtain a closed-form expression for the channel estimation variance. We then address the issue of superimposed training power allocation for complex Gaussian random (Rayleigh) channels. 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. Illustrative simulation examples are provided.

1. INTRODUCTION

Consider an 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 channel with discrete-time impulse response $\{\mathbf{h}(l)\}$. Then the symbol-rate, channel output vector is given by

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

The noisy measurements of $\mathbf{x}(n)$ are given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n). \qquad (2)$$

In superimposed training one takes

$$s(n) = b(n) + c(n), \qquad (3)$$

$\{b(n)\}$ 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 compared with conventional time-multiplexed training. Superimposed training-based approaches have been discussed in [1]-[3] and [4] for SISO (single-input single-output) and/or SIMO systems. Periodic superimposed training for channel estimation via first-order statistics for SISO and/or SIMO systems have been discussed in [1], [3], [4] and [5]. The formulations of [3] and [4] allow for the possibility of having an unknown dc offset at the receiver whereas [1] and [5] do not. A performance analysis (closed-form solution for the channel estimation variance) is performed in [3] under the assumption of zero (or known) dc offset.

Objectives and Contributions: In this paper, we consider a performance analysis of the approach of [4] to obtain a closed-form expression for the channel estimation variance. Unlike [3], our analysis is valid for any dc offset. We also address the issue of superimposed training power allocation for complex Gaussian random (Rayleigh) channels. The issue of power allocation for periodic superimposed training has not been addressed in [1], [3], [4] and [5].

This work was supported by NSF under Grant ECS-0424145 and the US Army Research Office under Grant DAAD19-01-1-0539.

Notation: Superscripts H, T and $\dagger$ denote the 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. $\mathrm{tr}(A)$ is the trace of the matrix A.

2. FIRST-ORDER STATISTICS-BASED SOLUTION OF [4]

Here we briefly review the approach of [4], introduce notation and present our underlying system model assumptions. Assumptions (**H1**)-(**H3**) are as in [4] whereas (**H4**) is not needed by the approach of [4].

Assume the following:

- (**H1**) The information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = \sigma_b^2$.
- (**H2**) 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 $\mathbf{m}$ is unknown. We will also use the notation $\mathbf{v}(n) = \tilde{\mathbf{v}}(n) + \mathbf{m}$.
- (**H3**) The superimposed training sequence $c(n) = c(n+mP)$ $\forall m, n$ is a non-random periodic sequence with period P. Let $\sigma_c^2 := (1/P)\sum_{n=1}^{P} |c(n)|^2$.
- (**H4**) Components of the channel coefficient $\mathbf{h}(l)$'s are assumed to be Gaussian random variables with zero mean and variance $\frac{1}{N(L+1)}$. We also assume that $\mathbf{h}_i(l)$ and $\mathbf{h}_m(k)$ are statistically independent if $k \neq l$ or $i \neq m$.

Assumption (**H4**) is used in training power allocation considerations; it is otherwise not essential to this paper. Under (**H4**), the received SNR is given by $(\sigma_b^2 + \sigma_c^2)/\sigma_v^2$.

By (**H3**), we have $c_m := \frac{1}{P}\sum_{n=0}^{P-1} c(n) e^{-j\alpha_m n}$,

$$c(n) = \sum_{m=0}^{P-1} c_m e^{j\alpha_m n} \;\; \forall n, \;\; \alpha_m := 2\pi m/P. \qquad (4)$$

The coefficients c_m's are known at the receiver since $\{c(n)\}$ is known. We have

$$E\{\mathbf{y}(n)\} = \sum_{m=0}^{P-1} \underbrace{\left[\sum_{l=0}^{L} c_m \mathbf{h}(l) e^{-j\alpha_m l}\right]}_{=:\mathbf{d}_m} e^{j\alpha_m n} + \mathbf{m}. \qquad (5)$$

In [4] $\mathbf{d}_m$ is estimated as $\hat{\mathbf{d}}_m = \dfrac{1}{T}\displaystyle\sum_{n=1}^{T} \mathbf{y}(n) e^{-j\alpha_m n}$. (6)

Define

$$\mathcal{H} := \begin{bmatrix} \mathbf{h}^H(0) & \mathbf{h}^H(1) & \cdots & \mathbf{h}^H(L) \end{bmatrix}^H, \qquad (7)$$

$$\mathcal{D} := \begin{bmatrix} \mathbf{d}_1^H & \mathbf{d}_2^H & \cdots & \mathbf{d}_{P-1}^H \end{bmatrix}^H, \tag{8}$$

$$\mathcal{C} := \underbrace{(\text{diag}\{c_1, c_2, \cdots, c_{P-1}\}\mathbf{V})}_{=:\mathcal{V}} \otimes I_N, \tag{9}$$

$$\mathbf{V} := \begin{bmatrix} 1 & e^{-j\alpha_1} & \cdots & e^{-j\alpha_1 L} \\ 1 & e^{-j\alpha_2} & \cdots & e^{-j\alpha_2 L} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\alpha_{P-1}} & \cdots & e^{-j\alpha_{P-1} L} \end{bmatrix}_{(P-1)\times(L+1)}. \tag{10}$$

From (5), it follows that

$$\mathcal{C}\mathcal{H} = \mathcal{D}. \tag{11}$$

It is shown in [4] that if $P-1 \geq L+1$ and α_i's are distinct, rank$(\mathcal{C}) = N(L+1)$; hence, we can determine $\mathbf{h}(l)$'s uniquely. Define $\hat{\mathcal{D}}$ as in (11) with $\mathbf{d}_m$'s replaced with $\hat{\mathbf{d}}_m$'s. Then we have the channel estimate

$$\hat{\mathcal{H}} = \mathcal{C}^\dagger \hat{\mathcal{D}} := (\mathcal{C}^H\mathcal{C})^{-1}\mathcal{C}^H\hat{\mathcal{D}}. \tag{12}$$

Define

$$\bar{c}(n) := \sigma_c^{-1}c(n), \quad \bar{c}_m := \sigma_c^{-1}c_m, \quad \bar{\mathcal{C}} := \sigma_c^{-1}\mathcal{C} \tag{13}$$

so that $(1/P)\sum_{n=1}^{P}|\bar{c}(n)|^2 = 1$. Then we may rewrite (12) as

$$\hat{\mathcal{H}} = \sigma_c^{-1}\bar{\mathcal{C}}^\dagger\hat{\mathcal{D}}. \tag{14}$$

Note that $\bar{\mathcal{C}}$ does not depend upon the training power σ_c^2.

3. PERFORMANCE ANALYSIS

In this section we present a performance analysis of the approach of [4] (briefly reviewed in Sec. 2). We provide closed-form expressions when the underlying channel is non-random and when (**H4**) holds true. Unlike [3], our analysis is applicable to the case of unknown dc offset $\mathbf{m}$.

Given the training sequence $c(n)$ and channel length $L+1$, $\mathcal{C}$ is known. By eq. (12),

$$\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}}|\mathcal{H}\} := E\{[\hat{\mathcal{H}} - E\{\hat{\mathcal{H}}|\mathcal{H}\}][\hat{\mathcal{H}} - E\{\hat{\mathcal{H}}|\mathcal{H}\}]^H|\mathcal{H}\}$$
$$= \mathcal{C}^\dagger \text{cov}\{\hat{\mathcal{D}}, \hat{\mathcal{D}}|\mathcal{H}\}\mathcal{C}^{\dagger H}. \tag{15}$$

We first calculate cov$\{\hat{\mathcal{D}}, \hat{\mathcal{D}}|\mathcal{H}\}$. Suppose that record length $T = KP$ for some positive integer K. By (1)-(5), it follows that

$$\mathbf{y}(n) = E\{\mathbf{y}(n)\} + \underbrace{\sum_{l=0}^{L}\mathbf{h}(l)b(n-l) + \tilde{\mathbf{v}}(n)}_{=:\tilde{\mathbf{x}}(n)}. \tag{16}$$

Using (6) and (16), we have

$$\hat{\mathbf{d}}_m = \frac{1}{KP}\sum_{n=1}^{KP}\left\{\sum_{\bar{m}=0}^{P-1}\mathbf{d}_{\bar{m}}e^{j\alpha_{\bar{m}}n} + \tilde{\mathbf{x}}(n)\right\}e^{-j\alpha_m n}$$
$$= \sum_{\bar{m}=0}^{P-1}\mathbf{d}_{\bar{m}}\left[\frac{1}{KP}\sum_{n=1}^{KP}e^{j(\alpha_{\bar{m}}-\alpha_m)n}\right] + \mathbf{v}_m \tag{17}$$

where

$$\mathbf{v}_m := \frac{1}{KP}\sum_{n=1}^{KP}\tilde{\mathbf{x}}(n)e^{-j\alpha_m n}. \tag{18}$$

Since, for any $K \geq 1$, we have

$$\frac{1}{KP}\sum_{n=1}^{KP}e^{j(\alpha_{\bar{m}}-\alpha_m)n} = \delta(m-\bar{m}), \tag{19}$$

we have

$$\hat{\mathbf{d}}_m = \mathbf{d}_m + \mathbf{v}_m. \tag{20}$$

Hence, by (**H1**) and (**H2**), we have

$$E\{\hat{\mathbf{d}}_m|\mathcal{H}\} = \mathbf{d}_m, \text{ hence } E\{\hat{\mathcal{H}}|\mathcal{H}\} = \mathcal{H} \text{ if } T = KP. \tag{21}$$

Define

$$\mathcal{M}_{mp} := E\{[\hat{\mathbf{d}}_m - \mathbf{d}_m][\hat{\mathbf{d}}_p - \mathbf{d}_p]^H|\mathcal{H}\} = E\{\mathbf{v}_m\mathbf{v}_p^H|\mathcal{H}\}$$
$$= \frac{1}{(KP)^2}\sum_{n_1=1}^{KP}\sum_{n_2=1}^{KP}\underbrace{E\{\tilde{\mathbf{x}}(n_1)\tilde{\mathbf{x}}^H(n_2)|\mathcal{H}\}}_{=:\mathbf{R}_{\tilde{x}}(n_1-n_2)}e^{-j\alpha_m n_1}e^{j\alpha_p n_2} \tag{22}$$

since $\tilde{\mathbf{x}}(n)$ defined in (16) is wide-sense stationary. Set $n_1 - n_2 = \tau$ in (22) to obtain

$$\mathcal{M}_{mp} = \frac{1}{KP}\sum_{\tau=1-KP}^{KP-1}\{\mathbf{R}_{\tilde{x}}(\tau)e^{-j\alpha_m\tau}\underbrace{[\frac{1}{KP}\sum_{n_2=1}^{KP}e^{j(\alpha_p-\alpha_m)n_2}]}_{\delta(p-m)}\}$$
$$= \frac{1}{KP}\mathcal{S}_{\tilde{x}}(\alpha_m)\delta(p-m) \tag{23}$$

where $\mathcal{S}_{\tilde{x}}(\alpha_m)$ is the power spectral density of $\{\tilde{x}(n)\}$ (conditioned on $\mathcal{H}$) at frequency α_m rad./sec., defined as

$$\mathcal{S}_{\tilde{x}}(\alpha_m) := \sum_\tau \mathbf{R}_{\tilde{x}}(\tau)e^{-j\alpha_m\tau} = \sigma_b^2\mathbf{H}(e^{j\alpha_m\tau})\mathbf{H}^H(e^{j\alpha_m\tau}) + \sigma_v^2 I_N \tag{24}$$

$$\mathbf{H}(e^{j\alpha}) := \sum_{l=0}^{L}\mathbf{h}(l)e^{-j\alpha l} \text{ and } \mathbf{H}^H(e^{j\alpha}) := \sum_{l=0}^{L}\mathbf{h}^H(l)e^{j\alpha l}. \tag{25}$$

Substitute (23) into (15) to obtain (we set $KP = T$)

$$\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}}|\mathcal{H}\} = \frac{1}{T}\mathcal{C}^\dagger[\text{block}-\text{diag}\{\mathcal{S}_{\tilde{x}}(\alpha_1), \mathcal{S}_{\tilde{x}}(\alpha_2),$$
$$\cdots, \mathcal{S}_{\tilde{x}}(\alpha_{P-1})\}]\mathcal{C}^{\dagger H}. \tag{26}$$

Invoking assumption (**H4**), we have

$$E_{\mathcal{H}}\{\mathcal{S}_{\tilde{x}}(\alpha)\} = \left(\sigma_b^2N^{-1} + \sigma_v^2\right)I_N \ \ \forall\,\alpha. \tag{27}$$

Hence, under (**H4**), it follows that

$$E_{\mathcal{H}}\{\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}}|\mathcal{H}\}\} = \frac{\sigma_b^2N^{-1} + \sigma_v^2}{T}\mathcal{C}^\dagger\mathcal{C}^{\dagger H} = \frac{\sigma_b^2N^{-1} + \sigma_v^2}{\sigma_c^2 T}\underbrace{\bar{\mathcal{C}}^\dagger\bar{\mathcal{C}}^{\dagger H}}_{=:\Gamma}. \tag{28}$$

Note that Γ in (28) is not a function of σ_c^2 or σ_v^2. Also, (28) holds true for all T for which $T = KP$, ($K > 0$ is an integer), or for "large" T.

The variance of the channel estimate will be defined as

$$\sigma_{\hat{\mathbf{h}}}^2 := E_{\mathcal{H}}\{E\{\|\hat{\mathcal{H}} - E\{\hat{\mathcal{H}}|\mathcal{H}\}\|^2|\mathcal{H}\}\} \tag{29}$$

$$= \text{tr}\left\{E_{\mathcal{H}}\{\text{cov}\{\hat{\mathcal{H}}, \hat{\mathcal{H}}|\mathcal{H}\}\}\right\} = \frac{\sigma_b^2N^{-1} + \sigma_v^2}{\sigma_c^2 T}\underbrace{\text{tr}\Gamma}_{=:\gamma}. \tag{30}$$

If the channel is non-random, then we may take trace of (26) as the variance of the channel estimation.

4. TRAINING POWER ALLOCATION

In this section we consider the issue of superimposed training power allocation for complex Gaussian random channels (assumption (**H4**)). Define the training power overhead β as (see (3))

$$\beta := \frac{\frac{1}{P}\sum_{n=1}^{P}|c(n)|^2}{\frac{1}{P}\sum_{n=1}^{P}E\{|s(n)|^2\}} = \frac{\sigma_c^2}{\sigma_b^2+\sigma_c^2}. \tag{31}$$

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. Let $\hat{\mathbf{h}}(l)$ denote the estimate of $\mathbf{h}(l)$ based on (12), i.e. $\hat{\mathbf{h}}(l)$ is the $(l+1)$th block component of $\hat{\mathcal{H}}$. We can rewrite (2) as

$$\mathbf{y}(n) = \sum_{l=0}^{L}\hat{\mathbf{h}}(l)s(n-l) + \sum_{l=0}^{L}[\mathbf{h}(l)-\hat{\mathbf{h}}(l)]s(n-l) + \tilde{\mathbf{v}}(n) + \mathbf{m} \tag{32}$$

Following [4], let the estimate of the noise mean be given

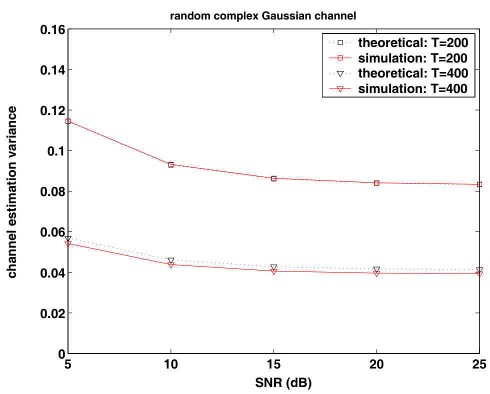

Figure 1. Channel estimation variance vs received SNR. Record length = 200 or 400 symbols. Results based on 100 Monte Carlo runs.

by

$$\hat{\mathbf{m}} := (1/T)\sum_{n=1}^{T}[\mathbf{y}(n) - \sum_{l=0}^{L}\hat{\mathbf{h}}(l)c(n-l)]. \tag{33}$$

Removing the estimated time-varying mean from the received data, define

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

$$+\sum_{l=0}^{L}[\mathbf{h}(l)-\hat{\mathbf{h}}(l)][b(n-l)+c(n-l)] + \tilde{\mathbf{v}}(n) + \mathbf{m} - \hat{\mathbf{m}}. \tag{34}$$

We will use the approximation (assuming that $\mathbf{m} \approx \hat{\mathbf{m}}$)

$$\tilde{\mathbf{y}}(n) \approx \underbrace{\sum_{l=0}^{L}\hat{\mathbf{h}}(l)b(n-l)}_{=:\mathbf{x}_s(n)}$$

$$+\underbrace{\sum_{l=0}^{L}[\mathbf{h}(l)-\hat{\mathbf{h}}(l)][b(n-l)+c(n-l)] + \tilde{\mathbf{v}}(n)}_{=:\mathbf{w}(n)}. \tag{35}$$

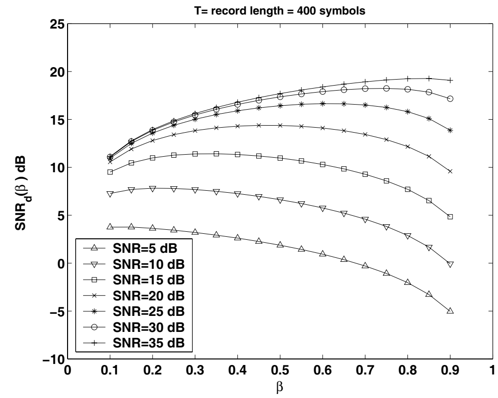

Figure 2. $SNR_d(\beta)$ (defined in (39)) vs β (defined in (31)).

When using $\hat{\mathbf{h}}(l)$ for equalization/detection, effective noise (as a first-order approximation) is $\mathbf{w}(n)$ whose covariance contains channel estimation error covariance as a component, which in turn will depend on β, and effective signal is $\mathbf{x}_s(n)$. An "optimum" value of the TIR α for the superimposed training method may be obtained by maximizing the SNR in (35). Using (**H1**), the signal power in (35) is given by

$$\sigma_{xs}^2 := E\{\|\mathbf{x}_s(n)\|^2\} = \sigma_b^2\sum_{l=0}^{L}E\{\|\hat{\mathbf{h}}(l)\|^2\}$$

$$= \sigma_b^2 E\{\|\hat{\mathcal{H}}\|^2\} = \sigma_b^2\,\mathrm{tr}\left\{E_{\mathcal{H}}\left\{E\{\hat{\mathcal{H}}\hat{\mathcal{H}}^H|\mathcal{H}\}\right\}\right\}$$

$$= \sigma_b^2\,\mathrm{tr}\left\{E_{\mathcal{H}}\left\{\mathrm{cov}\{\hat{\mathcal{H}},\hat{\mathcal{H}}|\mathcal{H}\} + \mathcal{H}\mathcal{H}^H\right\}\right\} = \sigma_b^2(\sigma_{\tilde{\mathbf{h}}}^2+1). \tag{36}$$

The noise power in (35) is given by

$$\sigma_w^2 := \frac{1}{P}\sum_{n=1}^{P}E\{\|\mathbf{w}(n)\|^2\} = \sigma_b^2\sigma_{\tilde{\mathbf{h}}}^2 + \sigma_v^2 + \frac{1}{P}\sum_{n=1}^{P}\sum_{l_1=0}^{L}$$

$$\cdot\sum_{l_2=0}^{L}E\{[\mathbf{h}(l_1)-\hat{\mathbf{h}}(l_1)]^H[\mathbf{h}(l_2)-\hat{\mathbf{h}}(l_2)]\}c^*(n-l_1)c(n-l_2). \tag{37}$$

The expected values in (37) can be obtained from traces of appropriate $N\times N$ submatrices of (28); thus, (37) can be computed. A simplification of (37) is possible if we use m-sequences (maximal length pseudo-random binary sequences) for $\{\bar{c}(n)\}$ in which case we have $\frac{1}{P}\sum_{n=1}^{P}\bar{c}^*(n-l_1)\bar{c}(n-l_2) = 1$ for $l_1 = l_2 \bmod P$, else $= \frac{-1}{P}$. For P "large," we may consider $\{\bar{c}(n)\}$ to be white. Under this approximation, we have

$$\sigma_w^2 \approx \sigma_{\tilde{\mathbf{h}}}^2(\sigma_b^2+\sigma_c^2) + \sigma_v^2. \tag{38}$$

Using (36) and (37) (or (38)), we obtain the SNR of (35) as a function of β as

$$\mathrm{SNR}_d(\beta) = \sigma_{xs}^2/\sigma_w^2. \tag{39}$$

Our objective is to maximize $\mathrm{SNR}_d(\beta)$ with respect to (w.r.t.) β under the constraint of a fixed received SNR $\mathcal{R}_{SNR}$, leading to the constraint

$$\sigma_b^2+\sigma_c^2 = \sigma_v^2\mathcal{R}_{SNR}. \tag{40}$$

With (40) holding true for a given received SNR $\mathcal{R}_{SNR}$, one can vary β and compute corresponding values of $\mathrm{SNR}_d(\beta)$ to pick an optimal β numerically.

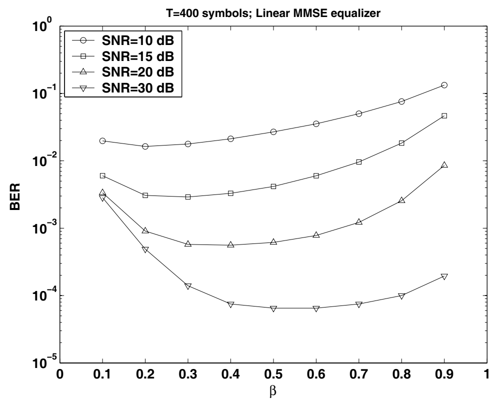

Figure 3. BER vs β (defined in (31)). Based on 500 Monte Carlo runs.

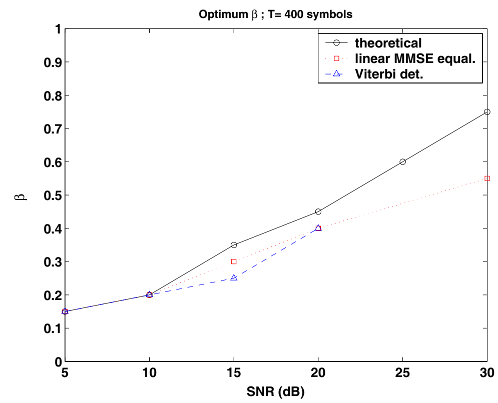

Figure 4. Optimal β versus received signal SNR, under three different criteria.

5. SIMULATION EXAMPLES

5.1. Example 1: Performance analysis

We took $N = 1$ and $L = 2$ in (1) with $h(l)$ mutually independent for all l, zero-mean complex-Gaussian with variance as in (**H4**). Additive noise was zero-mean complex white Gaussian to which we added an independent complex random mean m with independent real and imaginary parts, uniformly distributed over the interval $[0, a]$. The variable a was picked to achieve a specified dc-offset to signal ac-component (DCAC) power ratio defined as

$$\mathrm{DCAC} := E\{|m|^2\}[E\{|y(n) - v(n)|^2\}]^{-1}. \qquad (41)$$

The SNR refers to the energy per bit over one-sided noise spectral density with both information and superimposed training sequence counting toward the bit energy. Information sequence as well as superimposed training was binary. We took the superimposed training sequence period $P = 7$ in (**H3**); it is a scaled m-sequence (maximal length pseudo-random binary sequence) having a peak-to-average power ratio of one (the smallest possible). The average transmitted power in $c(n)$ (scaled binary) was 0.2 of the power in $b(n)$ – a small penalty in SNR, leading to the training-to-information sequence power ratio (TIR) of 0.2 .

Given the channel estimate (obtained via the method of Sec. 2) and the true channel at the i-th Monte Carlo run as $\hat{\mathbf{h}}^{(i)}(l)$ and $\mathbf{h}^{(i)}(l)$, respectively, the channel mean-square error (CMSE) is defined as

$$\mathrm{NCMSE} := \frac{1}{M_r}\sum_{i=1}^{M_r}\sum_{l=0}^{2}\|\hat{\mathbf{h}}^{(i)}(l) - \mathbf{h}^{(i)}(l)\|^2 \qquad (42)$$

where M_r is the number of Monte Carlo runs. The results averaged over 100 Monte Carlo for two different record lengths (T=200 or 400 bits) are shown in Fig. 1 where the simulations-based results are compared with the theoretical value $\sigma^2_{\hat{\mathbf{h}}}$ given by (30). Two cases were considered: DCAC ratios of 0.56 and 0. The theoretical results do not depend upon the DCAC ratio while in case of the simulations, the results are virtually indistinguishable (they do not show up in the Fig. 1). It is seen that the agreement between the theoretical and simulations results is quite good.

5.2. Example 2: Training power allocation

Here too we again consider the example of Sec. 5.1 (Example 1) except that the training-to-information sequence power ratio (TIR) is now varied to yield different values of β (see (31)). For a fixed received signal SNR ($= (\sigma_b^2 + \sigma_c^2)/\sigma_v^2$ under (**H4**)) with $\sigma_b^2 + \sigma_c^2 = 1$, we investigate the choice of β (defined in (31)) following Sec. 4. Our approach proposed in Sec. 4 is to choose β to maximize $\mathrm{SNR}_d(\beta)$ defined in (39). We maximize the theoretical expression for $\mathrm{SNR}_d(\beta)$ numerically by calculating it for different values of β with $\sigma_b^2 + \sigma_c^2 = 1$. Fig. 2 shows plots of $\mathrm{SNR}_d(\beta)$ versus β for a fixed T =400 symbols and varying SNR's. In this fig., β was varied in steps of 0.05 . It is seen that as received signal SNR increases, the optimum β increases too. Higher β implies that a higher fraction of transmitted power is allocated to training leading to more accurate channel estimates (with smaller estimation variance). Intuitively, for higher SNR's it pays to achieve more accurate channel estimates in order to achieve a lower effective noise power σ_w^2 (see (37)). On the other hand, when SNR is low (i.e. noise variance σ_v^2 is high), improving channel estimate does not have much effect on the effective noise power σ_w^2. The BER performance versus β based on simulation results (averaged over 500 Monte Carlo runs) is shown in Fig. 3 using a linear MMSE equalizer based on the channel estimate, equalizer length 11 and equalization delay 5, for a fixed record length of $T = 400$ symbols and varying SNR's. In Fig. 4 we compare the optimum values of β for a given received signal SNR for three cases: that maximizing theoretical $\mathrm{SNR}_d(\beta)$ (labeled "theoretical" in Fig. 4), that maximizing the BER based on linear MMSE (labeled "linear MMSE equal." in Fig. 4), and that maximizing the BER based on the Viterbi detector (labeled "Viterbi det." in Fig. 4). It is seen that the three curves follow the same trend, although they are not identical.

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.