PERFORMANCE ANALYSIS OF AN MIMO CHANNEL ESTIMATOR BASED ON SUPERIMPOSED TRAINING AND FIRST-ORDER STATISTICS

Jitendra K. Tugnait   *Shuangchi He*

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

ABSTRACT

Channel estimation for multiple-input multiple-output (MIMO) time-invariant channels using superimposed training is considered. Recently in [1], the first-order statistics-based approach of [11] was extended to multiuser systems (where semiblind versions using linear MMSE equalizers or Viterbi detectors were also presented). In this paper we present a performance analysis of the approach of [1] 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 for MIMO systems arising from spatial multiplexing of a single user signal. Illustrative simulation examples are provided.

1. INTRODUCTION

Accurate knowledge of the channel state information (CSI) of MIMO systems is a prerequisite for most MIMO physical layer approaches. In (conventional) training-based approaches, training sequences (known to the receiver), one per Tx antenna, 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.

Traditional approaches to single-input single-output (SISO) system CSI estimation have relied on time-multiplexed training sequences [10]. Various blind approaches for (some) MIMO systems may be found in [5], and [12]. Superimposed training-based approaches have been discussed in [6] and [8] for SISO systems. The UTRA specification for 3G systems [7] allows for a spread pilot (superimposed) sequence in the base station's common pilot channel, suitable for downlinks. Periodic superimposed training for channel estimation via first-order statistics for SISO systems have been discussed in [3], [9], [13], [11] and [14].

Consider an MIMO FIR (finite impulse response) linear channel with K inputs and N outputs. Let $\{s_k(n)\}$ denote k-th user's information sequence which is input to the MIMO channel with the k-th user's discrete-time impulse response $\{\mathbf{h}_k(l)\}$. 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

$$\mathbf{x}(n) := \sum_{k=1}^{K} \sum_{l=0}^{L} \mathbf{h}_k(l) s_k(n-l). \tag{1}$$

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

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

where $\mathbf{v}(n)$ is the additive Gaussian noise. [We will initially use the first-order statistics, i.e. $E\{\mathbf{y}(n)\}$, of the noisy data. To allow for mean-value ambiguity, we will take $E\{\mathbf{v}(n)\} = \mathbf{m}$, with $\mathbf{m}$ unknown; see also [11].] In **superimposed training**-based approaches for time-invariant systems, for user k, one takes

$$s_k(n) = b_k(n) + c_k(n) \tag{3}$$

in (1) where $\{b_k(n)\}$ is the information sequence and $\{c_k(n)\}$ is a non-random periodic training (pilot) sequence. Exploitation of the periodicity of $\{c_k(n)\}$ allows identification of the channel without allocating any explicit time slots for training, unlike traditional training methods. There is no loss in information rate. Unlike conventional training methods where one must know where the sequence begins and ends (synchronization), in superimposed training one can begin anywhere since it is always present (of course, proper phase synchronization is still required).

Objectives and Contributions: Recently in [1], the first-order statistics-based approach of [11] was extended to multiuser systems (where semiblind versions using linear MMSE equalizers or Viterbi detectors were also presented). In this paper we present a performance analysis of the approach of [1] to obtain a closed-form expression for the channel estimation variance. The analytical results are verified via simulations. We also address the issue of superimposed training power allocation for complex Gaussian random (Rayleigh) channels for MIMO systems arising from spatial multiplexing of a single user signal. 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 example is provided.

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.

2. FIRST-ORDER STATISTICS-BASED SOLUTION OF [1]

2.1. Multiuser (MIMO) Channels

Here we summarize some results from [1]. Consider K users with mutually independent information sequences $b_k(n)$ ($k = 1, 2, \cdots K$) with N measurements per symbol at the receiver (obtained by oversampling and/or antenna array). Using (1) and (3), the noise-free measurements at symbol rate are given by

$$\mathbf{x}(n) := \sum_{k=1}^{K} \sum_{l=0}^{L} \mathbf{h}_k(l)[b_k(n-l) + c_k(n-l)] \tag{4}$$

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

where $\{c_k(n)\}$ is the hidden pilot for the k-th user. The noisy measurements are given by (2).

The main idea is to pick user-specific training sequences so that the problem of channel estimation is decoupled across the various users – this allows us to use the single user superimposed training based approach outlined in [11]. Our approach is to assign distinct cycle frequencies of the periodic hidden pilots to distinct users. Suppose that for every user k, $\{c_k(n)\}$ is periodic with period $P = \tilde{P}K$ where $\tilde{P}$ is a positive integer. Then, in general,

$$c_k(n) = \sum_{m'=0}^{P-1} c_{m'k} e^{j(2\pi m'/P)n} \quad \forall n \tag{5}$$

where

$$c_{m'k} = \frac{1}{P}\sum_{n=0}^{P-1} c_k(n) e^{-j(2\pi m'/P)n}. \tag{6}$$

Pick $\{c_k(n)\}$ so that only $\tilde{P}$ coefficients (out of total P) $c_{m'k}$, associated with $\tilde{P}$ distinct frequencies, are nonzero. For instance, we may choose

$$c_k(n) = \sum_{m=0}^{\tilde{P}-1} c_{mk} e^{j(2\pi/P)[Km+k-1]n} \; \forall n \tag{7}$$

$$\alpha_{mk} := (2\pi/P)[Km + k - 1], \tag{8}$$

such that $c_{mk} \neq 0 \; \forall m, k$. Then

$$E\{\mathbf{y}(n)\} = \sum_{k=1}^{K}\sum_{m=0}^{\tilde{P}-1} \underbrace{\left[\sum_{l=0}^{L} c_{mk}\mathbf{h}_k(l)e^{-j\alpha_{mk}l}\right]}_{=:\mathbf{d}_{mk}} e^{j\alpha_{mk}n} + \mathbf{m}. \tag{9}$$

The model assumptions in [1] are as follows.

- (**H1**) The information sequences $\{b_k(n)\}$ are zero-mean, finite-alphabet, i.i.d. with $E\{|b_k(n)|^2\} = \sigma_{bk}^2$ and mutually independent for $k = 1, 2, \cdots, K$.
- (**H2**) The measurement noise $\{\mathbf{v}(n)\}$ in (2) is possibly **nonzero-mean** ($E\{\mathbf{v}(n)\} = \mathbf{m}$), white Gaussian, uncorrelated with $\{b_k(n)\}$, 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 sequences $c_k(n) = c_k(n + mP) \; \forall m, n$, are non-random periodic sequences with period P, satisfying (7) such that $c_{mk} \neq 0 \; \forall m, k$, and $\tilde{P}$ is an integer with $P = \tilde{P}K$.

2.2. First-Order Statistics-Based Solution of [1]

It follows from model assumptions that for $k_1 \neq k_2$, we have $\alpha_{m_1k_1} \neq \alpha_{m_2k_2}$ for any $m_1, m_2 \in \{0, 1, \cdots, \tilde{P} - 1\}$. Therefore, the sequence $E\{\mathbf{y}(n)\}$ is periodic in n with cycle frequencies [2] α_{mk}, $0 \leq m \leq \tilde{P} - 1$, $k = 1, 2, \cdots, K$, where α_{mk}'s are distinct. A consistent (mean-square (m.s.) sense) estimate $\hat{\mathbf{d}}_{mk}$ of $\mathbf{d}_{mk}$, for $\alpha_{mk} \neq 0$, follows as [2]

$$\hat{\mathbf{d}}_{mk} := \frac{1}{T}\sum_{n=0}^{T-1} \mathbf{y}(n) e^{-j\alpha_{mk}n}. \tag{10}$$

As $T \to \infty$, $\hat{\mathbf{d}}_{mk} \to \mathbf{d}_{mk}$ m.s. if $\alpha_{mk} \neq 0$ [2] and $\hat{\mathbf{d}}_{01} \to \mathbf{d}_{01} + \mathbf{m}$ m.s. since $\alpha_{mk} = 0$ iff $m = 0$ and $k = 1$.

Given $\mathbf{d}_{mk}$ for $1 \leq m \leq \tilde{P} - 1$, $k = 1, 2, \cdots, K$, we can (uniquely) estimate $\mathbf{h}_k(l)$'s if $\tilde{P} \geq L + 2$ and (**H3**) is true. Since $\mathbf{m}$ is unknown, we will omit the term corresponding to $\alpha_{mk} = 0$. Define

$$\mathcal{V} := \begin{bmatrix} 1 & e^{-j\alpha_{1k}} & \cdots & e^{-j\alpha_{1k}L} \\ 1 & e^{-j\alpha_{2k}} & \cdots & e^{-j\alpha_{2k}L} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\alpha_{(\tilde{P}-1)k}} & \cdots & e^{-j\alpha_{(\tilde{P}-1)k}L} \end{bmatrix}, \tag{11}$$

$$\tilde{\mathcal{C}}_k := \text{diag}\{c_{1k}, c_{2k}, \cdots, c_{(\tilde{P}-1)k}\}\mathcal{V}. \tag{12}$$

Omitting the terms corresponding to $m = 0$ and using the definition of $\mathbf{d}_{mk}$ from (9), we have

$$\underbrace{\left(\tilde{\mathcal{C}}_k \otimes I_N\right)}_{=:\mathcal{C}_k} \underbrace{\begin{bmatrix} \mathbf{h}_k(0) \\ \mathbf{h}_k(1) \\ \vdots \\ \mathbf{h}_k(L) \end{bmatrix}}_{=:\mathcal{H}_k} = \underbrace{\begin{bmatrix} \mathbf{d}_{1k} \\ \mathbf{d}_{2k} \\ \vdots \\ \mathbf{d}_{(\tilde{P}-1)k} \end{bmatrix}}_{=:\mathcal{D}_k} \tag{13}$$

for $k = 1, 2, \cdots, K$. Since α_{mk}'s are distinct and $c_{mk} \neq 0$ $\forall m \neq 0$, $\text{rank}(\tilde{\mathcal{C}}_k) = L + 1$ if $\tilde{P} \geq L + 2$ (see also [11]); hence, $\text{rank}(\mathcal{C}_k) = N(L + 1)$. Therefore, we can determine $\mathbf{h}_k(l)$'s uniquely. Define $\hat{\mathcal{D}}_k$ as in (13) with $\mathbf{d}_{mk}$'s replaced with $\hat{\mathbf{d}}_{mk}$'s. Then we have the channel estimate

$$\hat{\mathcal{H}}_k = \mathcal{C}_k^{\dagger}\hat{\mathcal{D}}_k = (\mathcal{C}_k^H\mathcal{C}_k)^{-1}\mathcal{C}_k^H\hat{\mathcal{D}}_k. \tag{14}$$

It is shown in [1] that under model assumptions, the channel estimator (14) is consistent in probability (i.p.) if $\tilde{P} \geq L + 2$.

3. PERFORMANCE ANALYSIS

In this section we present a performance analysis of the approach of [1] (briefly reviewed in Sec. 2). We provide closed-form expressions when the underlying channel is non-random and when (**H4**) holds true.

Define

$$\mathcal{H} := \left[\mathcal{H}_1^H \; \mathcal{H}_2^H \; \cdots \mathcal{H}_K^H\right]^H,$$

$$\mathbf{S}(\alpha) := \sigma_v^2 I_N + \sum_{k=1}^{K} \sigma_{bk}^2 \mathbf{H}_k(e^{j\alpha})\mathbf{H}_k^H(e^{j\alpha}), \tag{15}$$

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

Since $\lim_{T\to\infty} \hat{\mathbf{d}}_{mk} = \mathbf{d}_{mk}$ m.s. implies that $\lim_{T\to\infty} E\{\hat{\mathbf{d}}_{mk}\} = \mathbf{d}_{mk}$, it follows that

$$\lim_{T\to\infty} E\{\hat{\mathcal{H}}_k|\mathcal{H}\} = \mathcal{H}_k. \tag{16}$$

We now show that for large T

$$\text{cov}\{\hat{\mathcal{H}}_i, \hat{\mathcal{H}}_p|\mathcal{H}\} := E\{[\hat{\mathcal{H}}_i - E\{\hat{\mathcal{H}}_i|\mathcal{H}\}][\hat{\mathcal{H}}_p - E\{\hat{\mathcal{H}}_p|\mathcal{H}\}]^H|\mathcal{H}\}$$
$$= T^{-1}\mathcal{C}_i^{\dagger}\text{diag}\left\{\mathbf{S}(\alpha_{1i}),\; \mathbf{S}(\alpha_{2i}),\; \cdots,\; \mathbf{S}(\alpha_{(\tilde{P}-1)i})\right\}\mathcal{C}_i^{\dagger H}\delta(i-p). \tag{17}$$

Suppose that record length $T = MP$ for some positive integer M. By (1)-(9), it follows that

$$\mathbf{y}(n) = E\{\mathbf{y}(n)|\mathcal{H}\} + \underbrace{\sum_{k=1}^{K}\sum_{l=0}^{L} \mathbf{h}_k(l)b_k(n-l) + \tilde{\mathbf{v}}(n)}_{=:\tilde{\mathbf{x}}(n)}. \tag{18}$$

We have

$$\hat{\mathcal{D}}_k = (MP)^{-1} \sum_{n=0}^{MP-1} (\underbrace{\left[e^{j\alpha_{1k}n}\ e^{j\alpha_{2k}n}\ \cdots\ {}^{j\alpha_{(\tilde{P}-1)k}n}\right]^H}_{=:\mathbf{e}_k(n)}) \otimes \mathbf{y}(n). \tag{19}$$

By (18) and (19), it follows that

$$\mathrm{cov}\{\hat{\mathcal{D}}_i, \hat{\mathcal{D}}_p|\mathcal{H}\} = \frac{1}{(MP)^2} \sum_{n_1=0}^{MP-1} \sum_{n_2=0}^{MP-1} \left[\mathbf{e}_i(n_1)\mathbf{e}_p^H(n_2)\right]$$

$$\otimes \underbrace{E\{\tilde{\mathbf{x}}(n_1)\tilde{\mathbf{x}}^H(n_2)|\mathcal{H}\}}_{=:\mathbf{R}_{\tilde{x}}(n_1-n_2)} \tag{20}$$

since $\tilde{\mathbf{x}}(n)$ defined in (18) is wide-sense stationary. Set $n_1 - n_2 = \tau$ in (20) to obtain

$$\mathrm{cov}\{\hat{\mathcal{D}}_i, \hat{\mathcal{D}}_p|\mathcal{H}\} = \frac{1}{MP} \sum_{\tau=1-MP}^{MP-1}$$

$$\left[\frac{1}{MP} \sum_{n_2=\max(0,-\tau)}^{\min(MP-1, MP-1-\tau)} \mathbf{e}_i(\tau+n_2)\mathbf{e}_p^H(n_2)\right] \otimes \mathbf{R}_{\tilde{x}}(\tau). \tag{21}$$

We have

$$\frac{1}{P}\sum_{n=0}^{P} e^{j(\alpha_{m_1 k_1} - \alpha_{m_2 k_2})n} = \delta(k_1-k_2)\delta(m_1-m_2). \tag{22}$$

Using (22), it follows that

$$\frac{1}{MP}\sum_{n_2=0}^{MP-1} \mathbf{e}_i(\tau+n_2)\mathbf{e}_p^H(n_2) = \frac{1}{M}\sum_{m=1}^{M} \frac{1}{P}\sum_{l=0}^{P-1} \mathbf{e}_i(\tau+l)\mathbf{e}_p^H(l)$$

$$= \frac{1}{M}\sum_{m=1}^{M} \underbrace{\mathrm{diag}\{e^{-j\alpha_{1i}\tau},\ e^{-j\alpha_{2i}\tau},\ \cdots,\ e^{-j\alpha_{(\tilde{P}-1)i}\tau}\}}_{=:\Sigma_i} \delta(i-p). \tag{23}$$

Thus, from (21)-(23), it follows for "large" MP (note that $\mathbf{R}_{\tilde{x}}(\tau) = 0$ for $|\tau| > L$) that

$$\mathrm{cov}\{\hat{\mathcal{D}}_i, \hat{\mathcal{D}}_p|\mathcal{H}\} = \frac{1}{MP}\sum_{\tau=1-MP}^{MP-1} \Sigma_i \otimes \mathbf{R}_{\tilde{x}}(\tau)\delta(i-p). \tag{24}$$

Let $\mathbf{S}(\alpha)$ denote the power spectral density of $\{\tilde{\mathbf{x}}(n)\}$ (conditioned on $\mathcal{H}$) at frequency α rad./sec., defined as $\mathbf{S}(\alpha) := \sum_\tau \mathbf{R}_{\tilde{x}}(\tau)e^{-j\alpha\tau}$ and given by (15). Eqns. (15), (14) and (24) yield (we set $MP = T$) the expression (17).

For a non-random channel, the (conditional) variance of the channel estimate will be defined as

$$\sigma^2_{\hat{\mathbf{h}}|\mathcal{H}} := E\{\|\hat{\mathcal{H}} - E\{\hat{\mathcal{H}}|\mathcal{H}\}\|^2|\mathcal{H}\}$$

$$= \mathrm{tr}\left\{\mathrm{cov}\{\hat{\mathcal{H}},\hat{\mathcal{H}}|\mathcal{H}\}\right\} = \frac{1}{T}\sum_{i=1}^{K} \mathrm{tr}\left\{\mathcal{S}_i \mathcal{C}_i^{\dagger H}\mathcal{C}_i^{\dagger}\right\} \tag{25}$$

where

$$\mathcal{S}_i := \mathrm{diag}\left\{\mathbf{S}(\alpha_{1i}),\ \mathbf{S}(\alpha_{2i}),\ \cdots,\ \mathbf{S}(\alpha_{(\tilde{P}-1)i})\right\}. \tag{26}$$

3.1. Gaussian Random Channels

In certain cases a much simpler covariance expression can be obtained. Consider the following assumption on the channel.

(H4) Components of the channel coefficient $\mathbf{h}_k(l)$'s are assumed to be complex Gaussian random variables with zero mean and variance $\frac{1}{N(L+1)}$. We also assume that $[\mathbf{h}_k]_i(l_1)$ and $[\mathbf{h}_m]_j(l_2)$ are statistically independent if $l_1 \neq l_2$ or $k \neq m$ or $i \neq j$ where $[\mathbf{h}_k]_i(l)$ denotes the i-th component of $\mathbf{h}_k(l)$.

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

$$E_{\mathcal{H}}\{\mathbf{S}(\alpha)\} = (\sigma_v^2 + (\sum_{k=1}^{K}\sigma_{bk}^2)/N)I_N \ \ \forall\,\alpha. \tag{27}$$

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

$$E_{\mathcal{H}}\{\mathrm{cov}\{\hat{\mathcal{H}}_k, \hat{\mathcal{H}}_k|\mathcal{H}\}\} = \frac{\sigma_v^2 + (\sum_{k=1}^{K}\sigma_{bk}^2)/N}{T} \underbrace{\mathcal{C}_k^{\dagger}\mathcal{C}_k^{\dagger H}}_{=:\Gamma_k}. \tag{28}$$

Eqn. (28) holds true for all T for which $T = MP$, ($M > 0$ is an integer), or for "large" T.

The variance of the channel estimate for user k will be defined as

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

$$= \mathrm{tr}\left\{E_{\mathcal{H}}\{\mathrm{cov}\{\hat{\mathcal{H}}_k, \hat{\mathcal{H}}_k|\mathcal{H}\}\}\right\} = \frac{\sigma_v^2 + (\sum_{k=1}^{K}\sigma_{bk}^2)/N}{T}\mathrm{tr}\Gamma_k. \tag{30}$$

3.1.1. *White Superimposed Training*

A simplification of Γ_k in (28) is possible if we use m-sequences (maximal length pseudo-random binary sequences) for training. Pick $\tilde{P} = 2^n - 1$ for some integer n such that $\tilde{P} \geq L+2$. Let $\{\bar{c}_o(n)\}$ be an m-sequence of length $\tilde{P}$. Pick the superimposed training sequence $\{c_1(n)\}_{n=0}^{P-1}$ for user 1 as K repetitions of $\{\bar{c}_o(n)\}$ multiplied by a factor σ_{c1} so that $P^{-1}\sum_{n=0}^{P-1}|c_1(n)|^2 = \sigma_{c1}^2$. This choice satisfies (7)-(8) for $k=1$. Pick $\bar{c}_k(n) = \bar{c}_1(n)e^{j(2\pi/P)(k-1)n}$ for $k = 2, 3, \cdots, K$ and $c_k(n) = \sigma_{ck}\bar{c}_k(n)$. Then $\{c_k(n)\}_{n=0}^{P-1}$ satisfies (7)-(8) for $k \geq 2$. Furthermore, we also have $c_{mk} = (\sigma_{ck}/\sigma_{c1})c_{m1}\ \forall m$.

It then follows that $\frac{1}{\tilde{P}}\sum_{n=0}^{\tilde{P}-1}\bar{c}_1^*(n-l_1)\bar{c}_1(n-l_2) = 1$ for $l_1 = l_2 \bmod \tilde{P}$, else $= -\tilde{P}^{-1}$. We also have $c_{01} = \sigma_{c1}\tilde{P}^{-1}$. For $\tilde{P}$ "large," we may consider $\{\bar{c}_1(n)\}$ to be "exactly" white. Using (7)-(8), it easily follows that

$$\sum_{m=0}^{\tilde{P}-1}|c_{m1}|^2 e^{-j\alpha_{m1}l} = \sigma_{c1}^2\delta(l \bmod \tilde{P}) \tag{31}$$

leading to (for "large" $\tilde{P}$)

$$\sum_{m=1}^{\tilde{P}-1}|c_{m1}|^2 e^{-j\alpha_{m1}l} = \sigma_{c1}^2\delta(l \bmod \tilde{P}) - |c_{01}|^2$$

$$= \sigma_{c1}^2\left[\delta(l \bmod \tilde{P}) - \tilde{P}^{-2}\right] \approx \sigma_{c1}^2\delta(l \bmod \tilde{P}). \tag{32}$$

Similarly, for $k \geq 2$, we have $\frac{1}{\tilde{P}}\sum_{n=0}^{\tilde{P}-1}\bar{c}_k^*(n-l_1)\bar{c}_k(n-l_2) = e^{-j(2\pi/P)(k-1)(l_1-l_2)}$ for $l_1 = l_2 \bmod \tilde{P}$, else

$= -e^{-j(2\pi/P)(k-1)(l_1-l_2)}\tilde{P}^{-1}$. We also have $c_{0k} = (\sigma_{ck}/\sigma_{c1})c_{01} = \sigma_{ck}\tilde{P}^{-1}$. Then for $k \geq 2$, it follows that

$$\sum_{m=0}^{\tilde{P}-1} |c_{mk}|^2 e^{-j\alpha_{mk}l} = e^{-j(2\pi/P)(k-1)l}\sigma_{ck}^2\delta(l \bmod \tilde{P}) \quad (33)$$

leading to (for "large" $\tilde{P}$)

$$\sum_{m=1}^{\tilde{P}-1} |c_{mk}|^2 e^{-j\alpha_{mk}l} \approx e^{-j(2\pi/P)(k-1)l}\sigma_{ck}^2\delta(l \bmod \tilde{P}). \quad (34)$$

Therefore, for $k = 1, 2, \cdots, K$, we have

$$\Gamma_k = \left(\mathcal{C}_k^H \mathcal{C}_k\right)^{-1} \approx \sigma_{ck}^{-2} I_{N(L+1)}. \quad (35)$$

Therefore, for a white training signal, we have

$$\sigma_{\hat{\mathbf{h}}_k}^2 \approx \frac{\sigma_v^2 + (\sum_{k=1}^K \sigma_{bk}^2)/N}{\sigma_{ck}^2 T} \left(N(L+1)\right). \quad (36)$$

Finally we have

$$\sigma_{\hat{\mathbf{h}}}^2 := E_{\mathcal{H}}\{E\{\|\hat{\mathcal{H}} - E\{\hat{\mathcal{H}}|\mathcal{H}\}\|^2|\mathcal{H}\}\} \quad (37)$$

$$= \sum_{k=1}^K \sigma_{\hat{\mathbf{h}}_k}^2 \approx \frac{N(L+1)}{T}\left(\sigma_v^2 + (\sum_{k=1}^K \sigma_{bk}^2)/N\right)\left(\sum_{k=1}^K \sigma_{ck}^{-2}\right). \quad (38)$$

4. TRAINING POWER ALLOCATION FOR SPATIAL MULTIPLEXING

Consider spatial multiplexing [4]) where a single user is treated as K virtual users with a block of K symbols transmitted simultaneously from K transmitters, one symbol per transmitter. In this section we consider the issue of superimposed training power allocation under spatial multiplexing for complex Gaussian random channels (assumption (**H4**)). Under (**H4**), by symmetry, we have $\sigma_{bk}^2 = \sigma_b^2$ and $\sigma_{ck}^2 = \sigma_c^2$ $\forall k$: equal signal power for each user and similarly, equal training power for each user. Using the channel estimation variance expression developed earlier, we cast the power allocation problem as one of optimizing a signal-to-noise ratio (SNR) for equalizer design. Since the estimated channel would be used for equalizer design, we set up a corresponding model for the received signal in which the channel estimation error-related term and the additive noise act as effective additive noise and the information-sequence-driven estimated channel output acts as effective signal. This SNR is maximized under a power constraint on the sum of powers in the information and training sequences.

Define the training power overhead β as

$$\beta := \frac{\frac{1}{P}\sum_{n=1}^P |c_k(n)|^2}{\frac{1}{P}\sum_{n=1}^P E\{|s_k(n)|^2\}} = \frac{\sigma_c^2}{\sigma_b^2 + \sigma_c^2}. \quad (39)$$

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. We can rewrite (1) and (2) as

$$\mathbf{y}(n) = \sum_{k=1}^K \sum_{l=0}^L \hat{\mathbf{h}}_k(l)s_k(n-l) + \sum_{k=1}^K \sum_{l=0}^L [\mathbf{h}_k(l) - \hat{\mathbf{h}}_k(l)]s_k(n-l)$$

$$+\tilde{\mathbf{v}}(n) + \mathbf{m} \quad (40)$$

Following [1], let the estimate of the noise mean be given by

$$\hat{\mathbf{m}} := (1/T)\sum_{n=0}^{T-1}[\mathbf{y}(n) - \sum_{k=1}^K \sum_{i=0}^L \hat{\mathbf{h}}_k(i)c_k(n-i)]. \quad (41)$$

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

$$\tilde{\mathbf{y}}(n) := \mathbf{y}(n) - \sum_{k=1}^K \sum_{l=0}^L \hat{\mathbf{h}}_k(l)c_k(n-l) - \hat{\mathbf{m}}$$

$$= \sum_{k=1}^K \sum_{l=0}^L \hat{\mathbf{h}}_k(l)b_k(n-l) + \tilde{\mathbf{v}}(n) + \mathbf{m} - \hat{\mathbf{m}}$$

$$+ \sum_{k=1}^K \sum_{l=0}^L [\mathbf{h}_k(l) - \hat{\mathbf{h}}_k(l)][b_k(n-l) + c_k(n-l)]. \quad (42)$$

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

$$\tilde{\mathbf{y}}(n) \approx \underbrace{\sum_{k=1}^K \sum_{l=0}^L \hat{\mathbf{h}}_k(l)b_k(n-l)}_{=:\mathbf{x}_s(n)}$$

$$+ \underbrace{\sum_{k=1}^K \sum_{l=0}^L [\mathbf{h}_k(l) - \hat{\mathbf{h}}_k(l)][b_k(n-l) + c_k(n-l)] + \tilde{\mathbf{v}}(n)}_{=:\mathbf{w}(n)}. \quad (43)$$

When using $\hat{\mathbf{h}}_k(l)$ $(k = 1, 2, \cdots, K)$ 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 training overhead β for the superimposed training method may be obtained by maximizing the SNR in (43). The signal power in (43) is given by

$$\sigma_{xs}^2 := E\{\|\mathbf{x}_s(n)\|^2\} = \sigma_b^2 \sum_{k=1}^K \sum_{l=0}^L E\{\|\hat{\mathbf{h}}_k(l)\|^2\}$$

$$= \sigma_b^2 \sum_{k=1}^K E\{\|\hat{\mathcal{H}}_k\|^2\} = \sigma_b^2 \sum_{k=1}^K \mathrm{tr}\left\{E_{\mathcal{H}}\left\{E\{\hat{\mathcal{H}}_k\hat{\mathcal{H}}_k^H|\mathcal{H}\}\right\}\right\}$$

$$= \sigma_b^2 \sum_{k=1}^K \mathrm{tr}\left\{E_{\mathcal{H}}\left\{\mathrm{cov}\{\hat{\mathcal{H}}_k, \hat{\mathcal{H}}_k|\mathcal{H}\} + \mathcal{H}_k\mathcal{H}_k^H\right\}\right\}$$

$$= \sigma_b^2 \sum_{k=1}^K (\sigma_{\hat{\mathbf{h}}_k}^2 + 1). \quad (44)$$

The noise power in (43) is given by

$$\sigma_w^2 := \frac{1}{P}\sum_{n=1}^P E\{\|\mathbf{w}(n)\|^2\} = \sigma_b^2 \sum_{k=1}^K \sigma_{\hat{\mathbf{h}}_k}^2 + N\sigma_v^2 + \frac{1}{P}\sum_{n=1}^P \sum_{k_1=1}^K$$

$$\cdot \sum_{k_2=1}^K \sum_{l_1=0}^L \sum_{l_2=0}^L E\{[\mathbf{h}_{k_1}(l_1) - \hat{\mathbf{h}}_{k_2}(l_1)]^H[\mathbf{h}_{k_1}(l_2) - \hat{\mathbf{h}}_{k_1}(l_2)]\}$$

$$\cdot c_{k_1}^*(n-l_1)c_{k_2}(n-l_2). \tag{45}$$

The expected values in (45) can be obtained from traces of appropriate $N \times N$ submatrices of (28); thus, (45) can be computed.

Using (44) and (45), we obtain the SNR of (43) as a function of β (implicitly) as

$$\text{SNR}_d(\beta) = \frac{\sigma_{xs}^2}{\sigma_w^2}. \tag{46}$$

Our objective is to maximize $\text{SNR}_d(\beta)$ with respect to (w.r.t.) β under the constraint of a fixed transmitted power, leading to the constraint

$$\sigma_b^2 + \sigma_c^2 = q. \tag{47}$$

An explicit construction and optimization of $\text{SNR}_d(\beta)$ is given in Sec. 4.1 for white superimposed training.

4.1. White Superimposed Training

A simplification of (45) is possible if we use the sequences discussed in Sec. 3.1.1 and invoke whiteness. Under this approximation, (45) simplifies to

$$\sigma_w^2 \approx (\sigma_b^2 + \sigma_c^2)\sum_{k=1}^{K}\sigma_{\hat{\mathbf{h}}_k}^2 + N\sigma_v^2. \tag{48}$$

With β as in (39), using the constraint (47), we have $\sigma_c^2 = \beta q$ and $\sigma_b^2 = (1-\beta)q$. Thus, incorporating these constraint-carrying variables in (46) via (48), (30) and (44), we have (after some manipulations) an unconstrained cost

$$\text{SNR}_d(\beta) = \frac{f_1\beta^2 + f_2\beta + f_3}{g_1\beta + g_2} \tag{49}$$

where

$$f_1 = qK[K(L+1)-T], \quad f_2 = qK[T-2K(L+1)] - KN(L+1)\sigma_v^2,$$

$$f_3 = K(L+1)[qK + N\sigma_v^2], \tag{50}$$

$$g_1 = NT\sigma_v^2 - K^2q(L+1), \quad g_2 = K(L+1)[qK+N\sigma_v^2] = f_3. \tag{51}$$

Setting the first derivative of $\text{SNR}_d(\beta)$ w.r.t. β to zero, we obtain a quadratic equation in β

$$\beta^2 + 2\frac{g_2}{g_1}\beta + \frac{f_2g_2 - f_3g_1}{f_1g_1} = 0$$

with two roots, one of which is negative ($\beta < 0$), hence is excluded, and the other root is given by

$$\beta_o = \frac{g_2}{g_1}\left[-1 + \sqrt{1 + \frac{g_1(f_3g_1 - f_2g_2)}{g_2^2 f_1}}\right]. \tag{52}$$

Note that the above solution should be in the interval [0,1] to be acceptable (else the solution is either 0 or 1). [One can also vary β and compute corresponding values of $\text{SNR}_d(\beta)$ to pick an optimal β numerically.] For "large" T, it can be shown that $\frac{g_1(f_3g_1-f_2g_2)}{g_2^2f_1} > 0$ if the received SNR $\frac{Kq}{N\sigma_v^2} > 1$ (see (47)). In this case, $\beta_o > 0$.

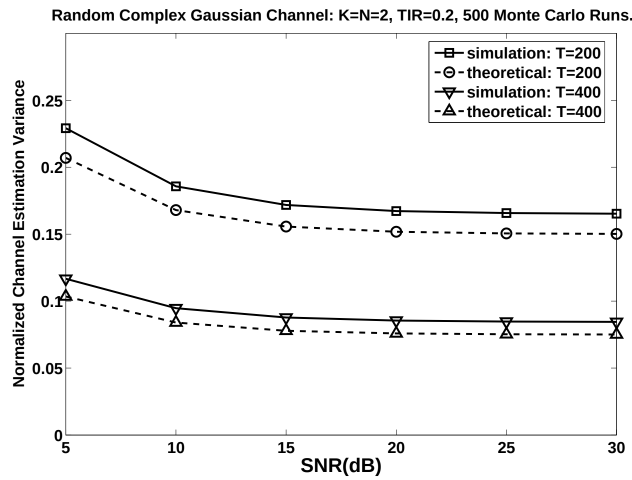

Figure 1. Normalized channel estimation variance vs received SNR. Record length = 200 or 400 symbols. Results based on 500 Monte Carlo runs. Solid/square or solid/down-triangle: simulation results; dashed/circle or dashed/up-triangle: theoretical expression.

5. SIMULATION EXAMPLES

5.1. Example 1: Performance analysis

We consider a two transmitter (a single user treated as two virtual users with alternate bits comprising two parallel data substreams transmitted simultaneously from the two transmitters: spatial multiplexing [4]) and a two receiver case with random frequency-selective Rayleigh fading channels leading to $K = N = 2$. We took $L = 2$ with the various components of $\mathbf{h}_k(l)$ mutually independent for all k and l, zero-mean Gaussian, with variance $1/(N(L+1))$ as in (**H4**). 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 was binary. We took $\tilde{P} = 7$ and $P = 14$ in (3). The sequence $\{\bar{c}_o(n)\}_{n=0}^{7}$ discussed in Sec. 3.1.1 is an m-sequence $\{1, -1, -1, 1, 1, 1, -1\}$. The two superimposed training sequences with period 14 were constructed from $\{\bar{c}_o(n)\}_{n=0}^{7}$ as discussed in Sec. 3.1.1. The average transmitted power in $c_k(n)$ (scaled binary) was 0.2 of the power in $b_k(n)$ $\forall k$– a small penalty in SNR. There was no loss in information rate. Moreover, $\sigma_{bk}^2 = \sigma_b^2$ $\forall k$.

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

$$\text{NCMSE} := \frac{\frac{1}{M_r}\sum_{i=1}^{M_r}\sum_{k=1}^{2}\sum_{l=0}^{2}\|\hat{\mathbf{h}}_k^{(i)}(l) - \mathbf{h}_k^{(i)}(l)\|^2}{2} \tag{53}$$

where M_r is the number of Monte Carlo runs and 2 in the denominator represents $\sum_{k=1}^{2}\sum_{l=0}^{2}E\{\|\mathbf{h}_k(l)\|^2\}$. The results averaged over 500 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 $\sigma_{\hat{\mathbf{h}}}^2/2$ where $\sigma_{\hat{\mathbf{h}}}^2$ is given by (38). [The factor 2 represents $\sum_{k=1}^{2}\sum_{l=0}^{2}E\{\|\mathbf{h}_k(l)\|^2\}$, as in (53).] It is seen that the agreement between the theoretical and simulations results is good.

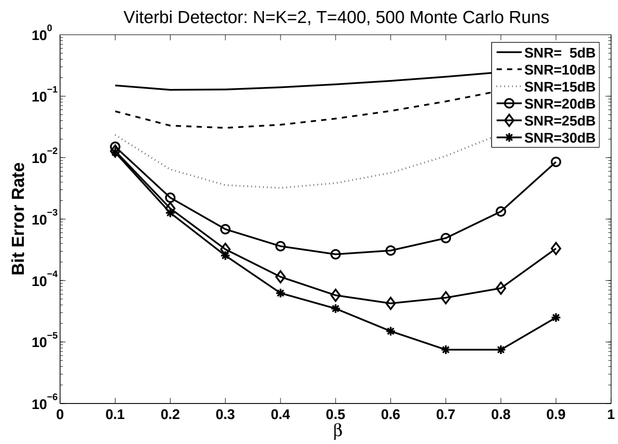

Figure 2. BER vs β (defined in (39)). Based on 500 Monte Carlo runs.

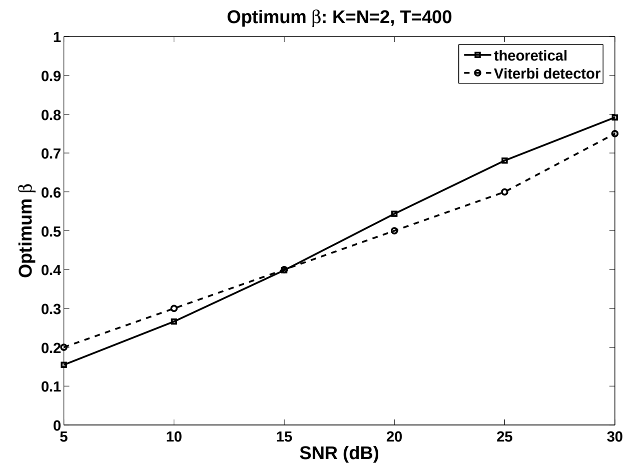

Figure 3. Optimal β versus received signal SNR, under two different criteria.

5.2. Example 2: Training power allocation

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 (39)). For a fixed received signal SNR ($= K(\sigma_b^2 + \sigma_c^2)/(N\sigma_v^2)$ under (**H4**)) with $\sigma_b^2 + \sigma_c^2 = 1$, we investigate the choice of β (defined in (39)) following Sec. 4. Our approach proposed in Sec. 4 is to choose β to maximize $\mathrm{SNR}_d(\beta)$ defined in (46).

We maximize the theoretical expression for $\mathrm{SNR}_d(\beta)$ using (52). The BER performance versus β based on simulation results (averaged over 500 Monte Carlo runs) is shown in Fig. 2 using a Viterbi detector, for a fixed record length of $T = 400$ symbols and varying SNR's. 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 (45)). 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. In Fig. 3 we compare the optimum values of β for a given received signal SNR for two cases: that maximizing theoretical $\mathrm{SNR}_d(\beta)$ (labeled "theoretical" in Fig. 3), and that maximizing the BER based on the Viterbi detector (labeled "Viterbi detector" in Fig. 3). It is seen that the two curves follow the same trend, although they are not identical.

6. CONCLUSIONS

We first considered a performance analysis of the approach of [1] to obtain a closed-form expression for the channel estimation variance for superimposed training-based MIMO channel estimation. We then addressed the issue of superimposed training power allocation for complex Gaussian random (Rayleigh) channels for MIMO systems arising from spatial multiplexing of a single user signal. All the results were illustrated via simulation examples involving frequency-selective Rayleigh fading.

REFERENCES

[1] X. Meng and J.K. Tugnait, "MIMO channel estimation using superimposed training," in *Proc. 2004 IEEE Intern. Conf. Commun.*, Paris, France, June 20-24, 2004.

[2] A.V. Dandawate and G.B. Giannakis, "Asymptotic theory of mixed time average and kth-order cyclic-moment and cumulant statistics," *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 216-232, Jan. 1995.

[3] 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.

[4] D. Gesbert et al, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. SAC-21, pp. 281-302, April 2003.

[5] G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong (Editors), *Signal Processing Advances in Wireless & Mobile Communications, Vol. 1: Trends in Channel Estimation and Equalization.* G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong (Eds.) Upper Saddle River, NJ: Prentice Hall, 2001.

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

[7] H. Holma and A. Toskala, Eds. *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, 2nd Ed. Wiley, 2002.

[8] 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.

[9] 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.

[10] J.G. Proakis, *Digital Communications, 4th ed.*. NY: McGraw-Hill, 2001.

[11] J.K. Tugnait and Weilin Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Letters*, vol. CL-7, pp. 413-415, Sept. 2003.

[12] J.K. Tugnait and Jinghong Ma, "Blind multiuser receiver for space-time coded CDMA signals in frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. WC-3, pp. 1770-1780, Sept. 2004.

[13] 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.

[14] J.K. Tugnait and X. Meng, "On superimposed training for channel estimation: Performance analysis, training power allocation and frame synchronization," *IEEE Trans. Signal Proc.*, to appear.