SEMI-BLIND CHANNEL ESTIMATION AND DETECTION USING SUPERIMPOSED TRAINING

Xiaohong Meng *Jitendra K. Tugnait*

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

ABSTRACT

Channel estimation for single-input multiple-output (SIMO) time-invariant or slowly time-varying channels is considered using superimposed training. 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. Two versions of a two-step approach are adopted where in the first step, following [11], we estimate the channel using only the first-order statistics of the data. Using the estimated channel from the first step, a linear MMSE equalizer and hard decisions, or a Viterbi detector, are used to estimate the information sequence. In the second step a deterministic maximum likelihood (DML) approach or an approximation to it, is used to iteratively estimate the SIMO channel and the information sequences sequentially. Illustrative computer simulation examples are presented where we compare the proposed approaches to the conventional (time-multiplexed) training based approach.

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 $\{\mathbf{h}(l)\}$. The vector channel may be the result of multiple receive antennas and/or oversampling at the receiver. Then the symbol-rate, channel output vector is given by

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

The noisy measurements of $\mathbf{x}(n)$ are given by ($\{\mathbf{v}(n)\}$ is possible nonzero-mean [11], temporally and spatially white, Gaussian)

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

A main objective in communications is to recover $s(n)$ given noisy $\{\mathbf{x}(n)\}$. In several approaches this requires knowledge of the channel impulse response [10], [8]. In training-based approach, $s(n) = c(n)$ = training sequence (known to the receiver) for (say) $n = 0, 1, \cdots, M_t - 1$ and $s(n)$ for $n > M_t - 1$ is the information sequence (unknown apriori to the receiver) [10], [8]. Therefore, given $c(n)$ and corresponding noisy $\mathbf{x}(n)$, one estimates the channel via least-squares and related approaches. For time-varying channels, one has to send training signal frequently and periodically to keep up with the changing channel. This wastes resources. An alternative is to estimate the channel based solely on noisy $\mathbf{x}(n)$ exploiting statistical and other properties of $\{s(n)\}$ [10], [8]. 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. This allows one to shorten the training period. Optimal placement and performance lower bounds for semi-blind approaches are in [1] and [2]. More recently a superimposed training based approach has been explored where one takes

$$s(n) = b(n) + c(n), \tag{3}$$

$\{b(n)\}$ 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. On the other hand, some useful power is wasted in superimposed training which could have otherwise been allocated to the information sequence. Superimposed training-based approaches have been discussed in [4], [5] and [7] for SISO systems. The UTRA specification for 3G systems [6] 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 [12] and [11]. In [3] performance bounds for training and superimposed training-based semiblind SISO channel estimation for time-varying flat fading channels have been discussed.

Objectives and Contributions: In this paper we extend the first-order statistics-based approach of [11] to semiblind versions using linear MMSE equalizers or Viterbi detectors. The first-order statistics-based approach views the information sequence as interference whereas in semiblind versions it is exploited to enhance channel estimation and information sequence detection.

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 [11]

Assume the following:

(H1) The information sequence $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = 1$.

(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.

(H3) The superimposed training sequence $c(n) = c(n+mP)$ $\forall m, n$ is a non-random periodic sequence with period P.

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

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

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

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

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

The sequence $E\{\mathbf{y}(n)\}$ is periodic with cycle frequencies α_m, $0 \le m \le P-1$. A mean-square (m.s.) consistent estimate $\hat{\mathbf{d}}_m$ of $\mathbf{d}_m$, for $\alpha_m \neq 0$, follows as

$$\hat{\mathbf{d}}_m = \frac{1}{T}\sum_{n=1}^{T} \mathbf{y}(n) e^{-j\alpha_m n}. \quad (6)$$

As $T \to \infty$, $\hat{\mathbf{d}}_m \to \mathbf{d}_m$ m.s. if $\alpha_m \neq 0$ and $\hat{\mathbf{d}}_0 \to \mathbf{d}_0 + \mathbf{m}$ m.s. if $\alpha_m = 0$.

It is established in [11] that given $\mathbf{d}_m$ for $1 \le m \le P-1$, we can (uniquely) estimate $\mathbf{h}(l)$'s if $P \ge L+2$, $\alpha_m \neq 0$, and $c_m \neq 0\ \forall m \neq 0$. Since $\mathbf{m}$ is unknown, we will omit the term $m = 0$ for further discussion. Define

$$\mathbf{V} := \begin{bmatrix} 1 & e^{-j\alpha_1} & \cdots & e^{-j\alpha_1 L} \\ 1 & e^{-j\alpha_2} & \cdots & e^{-j\alpha_2 L} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\alpha_{P-1}} & \cdots & e^{-j\alpha_{P-1} L} \end{bmatrix}_{(P-1)\times(L+1)}, \quad (7)$$

$$\mathcal{H} := \left[\ \mathbf{h}^H(0) \quad \mathbf{h}^H(1) \quad \cdots \quad \mathbf{h}^H(L)\ \right]^H, \quad (8)$$

$$\mathcal{D} := \left[\ \mathbf{d}_1^H \quad \mathbf{d}_2^H \quad \cdots \quad \mathbf{d}_{P-1}^H\ \right]^H, \quad (9)$$

$$\mathcal{C} := \underbrace{(\mathrm{diag}\{c_1, c_2, \cdots, c_{P-1}\}\mathbf{V})}_{=:\mathcal{V}} \otimes I_N. \quad (10)$$

Omitting the term $m = 0$ and using the definition of $\mathbf{d}_m$ from (5), it follows that

$$\mathcal{C}\mathcal{H} = \mathcal{D}. \quad (11)$$

It is shown in [11] that if $P-1 \ge L+1$ and α_i's are distinct, $\mathrm{rank}(\mathcal{C}) = N(L+1)$; hence, we can determine $\mathbf{h}(l)$'s uniquely. Define $\hat{\mathcal{D}}$ as in (9) with $\mathbf{d}_m$'s replaced with $\hat{\mathbf{d}}_m$'s. Then we have the channel estimate

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

.

3. ITERATIVE ENHANCEMENT

The first-order statistics-based approach of Sec. 2 views the information sequence as interference. Since the training and information sequences of a given user pass through identical channel, this fact can be exploited to enhance channel estimation performance via a semiblind approach. This is the objective of this section.

Suppose that we have collected $M-L$ samples of the observation $Y = [\mathbf{y}^T(M-1), \cdots, \mathbf{y}^T(L)]^T$. We then have the following linear model ($\tilde{\mathbf{v}}(n) := \mathbf{v}(n) - \mathbf{m}$)

$$Y = \mathcal{T}(\mathbf{s}) \underbrace{\begin{bmatrix} \mathbf{h}(0) \\ \vdots \\ \mathbf{h}(L) \end{bmatrix}}_{=:\mathbf{g}} + \underbrace{\begin{bmatrix} \tilde{\mathbf{v}}(M-1) \\ \vdots \\ \tilde{\mathbf{v}}(L) \end{bmatrix}}_{=:\tilde{V}} + \underbrace{\begin{bmatrix} \mathbf{m} \\ \vdots \\ \mathbf{m} \end{bmatrix}}_{=:\mathcal{M}} \quad (13)$$

where $V = \tilde{V} + \mathcal{M}$ is a column-vector consisting of samples of noise $\{\mathbf{v}(n)\}$, $\mathbf{g}$ is the vector of the channel parameters,

$$\mathcal{T}(\mathbf{s}) := \begin{bmatrix} s(M-1)I_N & \cdots & s(M-L-1)I_N \\ \mathit{Block} & \mathit{Hankel} & \mathit{Matrix} \\ s(L)I_N & \cdots & s(0)I_N \end{bmatrix} \quad (14)$$

and a block Hankel matrix has identical block entries on its block antidiagonals. In conventional training, we know (and have time-synchronization with) $s(n) = c(n)$ for $n = 0, 1, \cdots, M_t - 1$. In this case a least-squares (also maximum likelihood (ML) in white Gaussian noise) channel estimate is given by (set $M = M_t$ and $\mathbf{m} = 0$ in (13))

$$\widehat{\mathbf{g}} = \mathcal{T}^\dagger(\mathbf{c})Y = (\mathcal{T}^H(\mathbf{c})\mathcal{T}(\mathbf{c}))^{-1}\mathcal{T}^H(\mathbf{c})Y. \quad (15)$$

In our proposed iterative, superimposed training-based method we follow the following steps:

1) a) Use (12) to estimate the channel using the first-order (cyclostationary) statistics of the observations. Denote the channel estimate by $\widehat{\mathbf{g}}^{(1)}$ and $\widehat{\mathbf{h}}(l)$. In this method $\{c(n)\}$ is known and $\{b(n)\}$ is regarded as interference.
   b) Design a linear minimum mean-square error (LMMSE) equalizer of length L_e and equalization delay d using the estimated channel.
   c) Define (recall (1) and (2))

$$\tilde{\mathbf{y}}(n) := \mathbf{y}(n) - \sum_{i=0}^{L} \hat{\mathbf{h}}(i)c(n-i) - \hat{\mathbf{m}} \approx \sum_{i=0}^{L} \mathbf{h}(i)b(n-i) + \tilde{\mathbf{v}}(n) \quad (16)$$

$$\hat{\mathbf{m}} := (1/T)\sum_{n=1}^{T}[\mathbf{y}(n) - \sum_{i=0}^{L}\hat{\mathbf{h}}(i)c(n-i)]. \quad (17)$$

   Equalize the channel by applying the LMMSE equalizer to $\{\tilde{\mathbf{y}}(n)\}$ to estimate $\{b(n)\}$ as $\{\widehat{b}(n)\}$. Quantize $\{\widehat{b}(n)\}$ into $\{\tilde{b}(n)\}$ with the knowledge of the symbol alphabet (hard decisions).

2) a) Substitute $\tilde{s}(n) = \tilde{b}(n) + c(n)$ for $s(n)$ in (1) and use the corresponding formulation in (13) to estimate the channel $\mathbf{g}$ and mean $\mathbf{m}$ as ($\tilde{Y} := Y - \hat{\mathcal{M}}$)

$$\widehat{\mathbf{g}}^{(2)} = \mathcal{T}^\dagger(\tilde{\mathbf{s}})\tilde{Y} = (\mathcal{T}^H(\tilde{\mathbf{s}})\mathcal{T}(\tilde{\mathbf{s}}))^{-1}\mathcal{T}^H(\tilde{\mathbf{s}})\tilde{Y}, \quad (18)$$

$$\hat{\mathbf{m}}^{(2)} := (1/T)\sum_{n=1}^{T}[\mathbf{y}(n) - \sum_{i=0}^{L}\hat{\mathbf{h}}^{(2)}(i)c(n-i)]. \quad (19)$$

   b) Design a linear MMSE equalizer of length L_e and equalization delay d using the estimated channel $\widehat{\mathbf{g}}^{(2)}$ as in Step 1b.
   c) Repeat Step 1c using the results of Steps 2a, 2b.

3) Step 2 provides one iteration of our proposed iterative method. Repeat a few times if so desired.

Note that in Step 1a, the information sequence is treated as interference whereas in Step 2, it is exploited (along with the superimposed training) to further improve the results, leading to a semi-blind approach.

3.1. Simulation Results: LMMSE Equalizer

The results of a simulation example are shown in Figs. 1–2 for a random frequency-selective Rayleigh fading channel. We took $N = 1$ and $L = 2$ in (1) with $h(l)$ complex-valued (independent real and imaginary parts), mutually independent for all l, zero-mean unit variance Gaussian. Additive noise was zero-mean complex white Gaussian. 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**). The average transmitted power in $c(n)$ (scaled binary) was 0.2 of the power in $b(n)$ – a small penalty in SNR. There was no loss in information rate. Linear MMSE equalizer of length 11 bits and equalization delay of 5 bits was used throughout. The normalized channel mean-square error (NCMSE) is defined (before averaging over runs) as

$$\mathrm{NCMSE} := \left[\sum_{l=0}^{2} \|\widehat{\mathbf{h}}(l) - \mathbf{h}(l)\|^2\right] \left[\sum_{l=0}^{2} \|\mathbf{h}(l)\|^2\right]^{-1}. \quad (20)$$

We also implemented a conventional (time-multiplexed) training-based approach where the first 30 bits (± 1) were reserved for training and the remaining bits were information bits. We see that (Figs. 1 and 2) iterated enhancement is competitive with conventional training at lower SNR's (the "practical range"). Compared to conventional training there is no loss in information rate when superimposed training is used. Furthermore, we see that (Fig. 2) although the channel estimation errors may be much lower for superimposed training with iterated enhancement, because of SNR penalty (power wasted in superimposed training), this channel accuracy advantage does not necessarily translate into a large BER advantage.

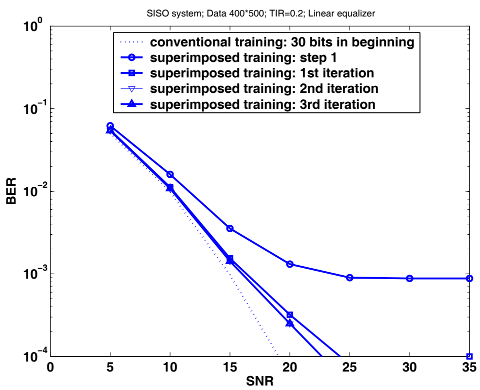

Figure 1. BER: circle: estimate channel using superimposed training and then design a linear MMSE equalizer; square: first iteration specified by Step 2 (Sec. 3); triangle down: second iteration specified by Step 2; triangle up: third iteration specified by Step 2; dots: estimate channel using conventional time-multiplexed training of length 30 bits and then design a linear MMSE equalizer. Training-to-information symbol power ratio =0.2 (-7 dB). Record length = 400 bits. Results based on 500 Monte Carlo runs.

4. DETERMINISTIC MAXIMUM LIKELIHOOD (DML) APPROACH

In this section we consider joint channel and information sequence estimation via an iterative DML approach. The main objective is as in Sec. 3, namely, to consider a semi-blind approach where the information sequence is not regarded as interference. Unlike the approach of Sec. 3 where convergence to a desired solution (a local maximum of DML function) is not guaranteed, here we have guaranteed convergence to a local maximum. Furthermore, if we initialize with our superimposed training-based solution, one is guaranteed the global extremum (minimum error probability sequence estimator) if the superimposed training-based solution is "good."

Consider (1), (2) and (13). Under the assumption of white Gaussian measurement noise, consider the joint estimators

$$\{\widehat{\mathbf{g}}, \widehat{\mathbf{s}}, \widehat{\mathbf{m}}\} = \arg\left\{\min_{\mathbf{g},\mathbf{s},\mathbf{m}} ||Y - \mathcal{T}(\mathbf{s})\mathbf{g} - \mathcal{M}||^2\right\} \quad (21)$$

where

$$\mathbf{s} := [s(M-1), s(M-2), \cdots, s(0)]^T \quad (22)$$

and $\widehat{\mathbf{s}}$ is the estimate of $\mathbf{s}$. In the above we have followed a deterministic ML (DML) approach assuming no statistical model for the input sequences $\{s(n)\}$. Under white Gaussian noise assumption, the DML estimators are obtained by the nonlinear least-squares optimization (21). Fortunately, the observation vector Y is linear in both the channel and the input parameters individually. In particular, we have

$$Y = \mathcal{T}(\mathbf{s})\mathbf{g} + \underbrace{\tilde{V} + \mathcal{M}}_{V} = \mathcal{C}(\mathbf{g})\mathbf{s} + \tilde{V} + \mathcal{M} \quad (23)$$

where

$$\mathcal{C}(\mathbf{g}) = \begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(L) & & \\ & \ddots & & \ddots & \\ & & \mathbf{h}(0) & \cdots & \mathbf{h}(L) \end{bmatrix} \quad (24)$$

is the the so-called filtering matrix. We therefore have a separable nonlinear least-squares problem that can be solved sequentially as (joint optimization w.r.t. $\mathbf{g}, \mathbf{m}$ can be further "separated")

$$\begin{aligned} \{\widehat{\mathbf{g}}, \widehat{\mathbf{s}}, \widehat{\mathbf{m}}\} &= \arg\min_{\mathbf{s}}\{\min_{\mathbf{g},\mathbf{m}} ||Y - \mathcal{T}(\mathbf{s})\mathbf{g} - \mathcal{M}||^2\} \quad (25) \\ &= \arg\min_{\mathbf{g},\mathbf{m}}\{\min_{\mathbf{s}} ||Y - \mathcal{C}(\mathbf{g})\mathbf{s} - \mathcal{M}||^2\}. \quad (26) \end{aligned}$$

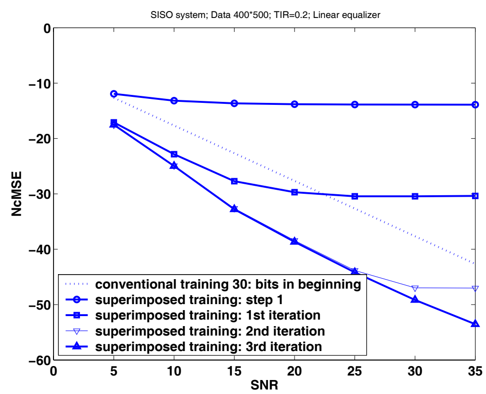

Figure 2. As in Fig. 1 except that NCMSE (normalized channel mean-square error) is shown.

The finite alphabet properties of the information sequences can also be incorporated into the deterministic maximum likelihood methods. These algorithms, first proposed by Seshadri [9] for SISO systems, iterate between estimates of the channel and the input sequences. At iteration k, with an initial guess of the channel $\mathbf{g}^{(k)}$ and the mean $\mathbf{m}^{(k)}$, the algorithm estimates the input sequence $\mathbf{s}^{(k)}$ and the channel $\mathbf{g}^{(k+1)}$ and mean $\mathbf{m}^{(k+1)}$ for the next iteration by

$$\begin{aligned} \mathbf{s}^{(k)} &= \arg\min_{\mathbf{s}\in\mathcal{S}} ||Y - \mathcal{C}(\mathbf{g}^{(k)})\mathbf{s} - \mathcal{M}^{(k)}||^2, \quad (27) \\ \mathbf{g}^{(k+1)} &= \arg\min_{\mathbf{g}} ||Y - \mathcal{T}(\mathbf{s}^{(k)})\mathbf{g} - \mathcal{M}^{(k)}||^2, \quad (28) \\ \mathbf{m}^{(k+1)} &= \arg\min_{\mathbf{m}} ||Y - \mathcal{T}(\mathbf{s}^{(k)})\mathbf{g}^{(k+1)} - \mathcal{M}||^2, \quad (29) \end{aligned}$$

where $\mathcal{S}$ is the (discrete) domain of $\mathbf{s}$. The optimizations in (28) and (29) are linear least squares problems whereas the the optimization in (27) can be achieved by using the Viterbi algorithm (VA) [8]. Since the above iterative procedure involving (27), (28) and (29) decreases the cost at every iteration, one achieves a local minimum of the nonlinear least-squares cost (local maximum of DML function).

We now summarize our DML approach:

1) a) Use (12) to estimate the channel using the first-order (cyclostationary) statistics of the observations. Denote the channel estimates by $\widehat{\mathbf{g}}^{(1)}$ and $\widehat{\mathbf{h}}(l)$. Estimate the mean via (17) as $\widehat{\mathbf{m}}^{(1)}$.
   b) Design a Viterbi sequence detector to estimate $\{s(n)\}$ as $\{\tilde{s}(n)\}$ using the estimated channel $\widehat{\mathbf{g}}^{(1)}$, mean $\widehat{\mathbf{m}}^{(1)}$ and cost (27) with $k = 1$. [Note that knowledge of $\{c(n)\}$ is used in $s(n) = b(n) + c(n)$, therefore, we are in essence estimating $b(n)$ in the Viterbi detector.]
2) a) Substitute $\tilde{s}(n)$ for $s(n)$ in (1) and use the corresponding formulation in (13) to estimate the channel $\mathbf{g}$ and mean $\mathbf{m}$ as

$$\widehat{\mathbf{g}}^{(2)} = \mathcal{T}^{\dagger}(\tilde{\mathbf{s}}) \left[Y - \widehat{\mathcal{M}}^{(1)} \right], \tag{30}$$

$$\widehat{\mathbf{m}}^{(2)} := (1/(M-L)) \sum_{n=L}^{M-1} [\mathbf{y}(n) - \sum_{i=0}^{L} \widehat{\mathbf{h}}^{(2)}(i)\tilde{s}(n-i)]. \tag{31}$$

   b) Design a Viterbi sequence detector using the estimated channel $\widehat{\mathbf{g}}^{(2)}$, mean $\widehat{\mathbf{m}}^{(2)}$ and cost (27) with $k = 2$, as in Step 1b.
3) Step 2 provides one iteration of (27)-(28). Repeat a few times if so desired.

4.1. Simulation Results: Viterbi Algorithm (VA)

We now repeat the example of Sec. 3.1 but use the iterative DML approach. [The various parameters are as in the example of Sec. 3.1.] The results corresponding to Figs. 1–2 of Sec. 3.1 are now shown in Figs. 3–4. The results based on VA are superior to that based on linear MMSE equalizers. The comments made regarding Figs. 1–2 apply to Figs. 3–4 also. [In Figs. 3–4 we display results for two types of conventional training sequences: 30 bits in the beginning or 30 bits in the middle; for time-invariant systems, one does not expect any differences.]

5. CONCLUSIONS

Approach of [11] to SIMO channel estimation using superimposed training sequences (hidden pilots) and first-order statistics was extended to semiblind versions thereof. The results were illustrated via a simulation example involving frequency-selective Rayleigh fading. The proposed methods are competitive with the conventional training method without incurring any information rate loss.

REFERENCES

[1] E. de Carvalho and D.T.M. Slock, "Semi-blind Methods for FIR Multichannel Estimation," Chapter 7 in *Signal Processing Advances in Wireless & Mobile Communications, vol. I*, G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong (Eds.) Upper Saddle River, NJ: Prentice Hall, 2001.

[2] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Proc.*, vol. 50, pp. 3055-3069, Dec. 2002.

[3] M. Dong L. Tong and B.M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," in *Proc. 2003 IEEE Workshop on Signal Processing Advances in Wireless Commun.*, pp. 472-476, Rome, Italy, June 15-18, 2003.

[4] 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.

[5] 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.

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

[7] 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.

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

[9] N. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," *IEEE Trans. Commun.*, vol. COM-42, pp. 1000-1011, March 1994.

[10] J.K. Tugnait, L. Tong and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Proc. Mag.*, vol. 17, pp. 16-28, May 2000.

[11] 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.

[12] 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.

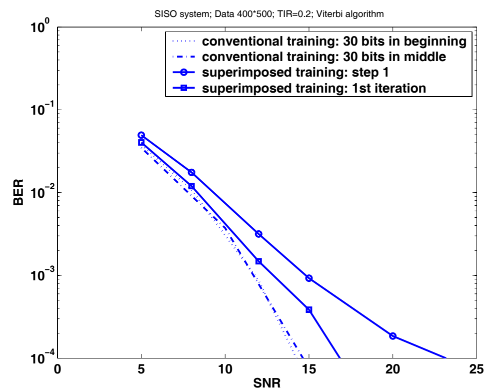

Figure 3. BER: circle: estimate channel using superimposed training and then design a Viterbi detector; square: first iteration specified by Step 2 (Sec. 4); dots: estimate channel using conventional time-multiplexed training of length 30 bits in the beginning and then design a Viterbi detector; dot-dashed: estimate channel using conventional time-multiplexed training of length 30 bits in the middle and then design a Viterbi detector. Training-to-information symbol power ratio =0.2 (-7 dB). Record length = 400 bits. Results based on 500 Monte Carlo runs.

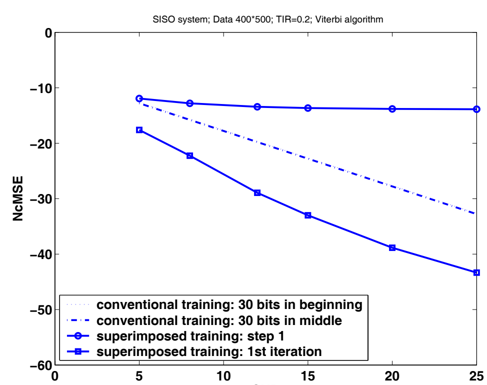

Figure 4. As in Fig. 3 except that NCMSE (normalized channel mean-square error) is shown.