

Semi-Blind Channel Estimation for Frequency-Selective MIMO Systems

Tianbin Wo and Peter Adam Hoeher

Faculty of Engineering, University of Kiel

Kaiserstr. 2, D-24143 Kiel, Germany

Phone/Fax: +49 431 880 6127/6128

Email: {wtb,ph}@tf.uni-kiel.de

Abstract—The task of channel estimation in a frequency-selective *multi-input multi-output* (MIMO) system presents a real challenge for conventional training-based solutions, since the huge number of channel coefficients demand a huge number of training symbols, which will significantly reduce the system bandwidth-efficiency. As a possible countermeasure, we propose a *semi-blind channel estimation* algorithm, which can learn channel coefficients accurately when only a short training is available. The main idea is to use blind data/channel estimation algorithms to improve the quality of the channel knowledge obtained from the short training. With the proposed semi-blind algorithm, the required training length will no longer be unaffordable for frequency-selective MIMO systems.

I. INTRODUCTION

For high-rate data transmission, MIMO systems attract increasing attention due to promising capacity gains [1][2]. If the delay spread is not significantly shorter than the symbol duration, the transmission channels are frequency-selective. Moreover, multiple data sequences arrive at individual receive antennas simultaneously and interfere with each other. Therefore, a MIMO channel equalizer is not only responsible for removing inter-symbol interference, but also for removing multiple-access interference. In order to accomplish this double duty, the equalizer often demands perfect channel knowledge.

In *single-input single-output* (SISO) systems, a training sequence is often inserted into the data stream for the purpose of *channel estimation* (CE). Because training-based algorithms have low computational complexity and good robustness in noisy environments, they are very popular in today's digital communication systems. However, if we directly apply purely training-based channel estimation schemes to MIMO systems, the situation becomes quite different. Due to inter-symbol interference and multiple-access interference, the required training length per transmit (Tx) antenna will be proportional to the product of the channel impulse response length and the number of Tx antennas. This fact presents a fundamental challenge for training-based CE schemes, especially when the system employs many Tx antennas and experiences a delay spread channel. The large amount of training symbols required for reliable channel estimation will significantly reduce the system bandwidth-efficiency. Furthermore, for a fast fading channel, it is also possible that the needed training length will spend a complete channel coherence interval, then there will be

no time left for data transmission before the channel changes. This situation should be definitely circumvented.

An intuitive countermeasure would be shortening the length of training. But, to make this practically attractive, a reduction of training length should not cause a considerable system performance degradation. As a matter of fact, all transmitted data symbols are actually carrying the same channel information as the training symbols, within the same channel coherence interval. If this channel information can also be exploited, then the required training length can be largely reduced. We propose a MIMO channel estimation algorithm, which can learn channel coefficients accurately when only a small amount of training is available. This algorithm is a combination of training-based channel estimation algorithms and blind equalization techniques, and is thereby termed as *semi-blind channel estimation* (SBCE).

Throughout this paper, the complex baseband notation is used. Concerning vector/matrix operations, we use $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^\dagger$ to denote transpose, complex conjugate, complex conjugate and transpose, and Moore-Penrose left/right pseudo inverse, respectively. A bold lower case letter marks a column vector, and a bold capital letter denotes a matrix.

II. CHANNEL MODEL

The equivalent discrete-time channel model of the MIMO system, which comprises pulse shaping, physical channel, receive filtering and baud-rate sampling, can be written as

$$y_i[k] = \sum_{j=1}^{N_T} \sum_{l=0}^L h_l^{ij}[k] \cdot x_j[k-l] + n_i[k], \quad (1)$$

where $(0 \leq k \leq K-1)$, k is the time index, and K is the block length. L denotes the effective memory length of all subchannels¹, which is assumed to be known at receiver side. $\mathbf{h}^{ij}[k] = [h_0^{ij}[k], h_1^{ij}[k], \dots, h_L^{ij}[k]]^T$ denotes the time-varying channel vector representing the subchannel between the j^{th} transmit antenna and the i^{th} receive antenna. $x_j[k]$ represents the k^{th} data symbol from the j^{th} transmit antenna, and $n_i[k]$ denotes the k^{th} additive white Gaussian noise sample added to the i^{th} channel output. Fig. 1 shows a block diagram of this channel model.

This work has been supported by the German Research Foundation (DFG) under contract no. HO 2226/6-1.

¹Here, we assume that all subchannels have the same effective memory length. However, this assumption does not indicate any restriction to the algorithm proposed in this paper.

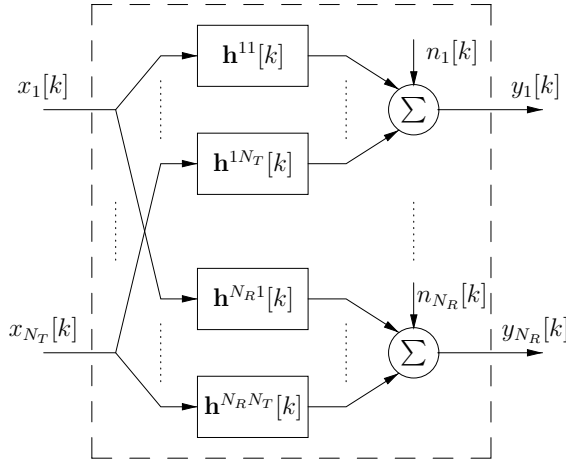


Fig. 1. Equivalent discrete-time MIMO channel model

Assuming block fading, (1) can be written in matrix form as

$$\mathbf{Y} = \mathbf{H} \cdot \mathbf{X} + \mathbf{N}, \quad (2)$$

where $\mathbf{Y}_{N_R \times K}$ is the channel output matrix, $\mathbf{H}_{N_R \times N_T(L+1)}$ is the channel matrix, $\mathbf{X}_{N_T(L+1) \times K}$ is the data matrix, and $\mathbf{N}_{N_R \times K}$ is the noise matrix. In the following, we will always use (2) instead of (1).

III. CHANNEL ESTIMATION

A. Joint Least Squares Channel Estimation

For SISO systems, *least-squares channel estimation* (LSCE) is optimal w.r.t. minimizing the squared channel estimation error, given that the data symbols are known during an observation interval [3]. For the MIMO case, if the training sequences of all Tx antennas overlap with each other for a time interval of sufficient length, LSCE will also be possible and optimal, and is then usually called *joint least squares channel estimation* (JLSCE) [4].

Given the channel model in (2), JLSCE can be expressed as

$$\hat{\mathbf{H}} = \mathbf{Y} \cdot \mathbf{X}_t^H \cdot (\mathbf{X}_t \mathbf{X}_t^H)^{-1} = \mathbf{Y} \cdot \mathbf{X}_t^\dagger, \quad (3)$$

where \mathbf{X}_t denotes the symbol matrix formed by training. The corresponding normalized (w.r.t. the number of Rx antennas)² *mean-squared error* (MSE) is defined as

$$\begin{aligned} MSE &\triangleq \mathbb{E} \left\{ \left\| \hat{\mathbf{H}} - \mathbf{H} \right\|_F^2 \right\} / N_R \\ &= \sigma_n^2 \cdot \text{tr} \left\{ (\mathbf{X}_t \mathbf{X}_t^H)^{-1} \right\}, \end{aligned} \quad (4)$$

where σ_n^2 denotes the variance of the additive noise, and $\text{tr}(\mathbf{A}) = \sum_{i=1}^m a_{ii}$ denotes the trace of matrix \mathbf{A} . For simplicity, we herein define

$$\text{Tri}(\mathbf{A}) \triangleq \text{tr} \left\{ (\mathbf{A} \mathbf{A}^H)^{-1} \right\}. \quad (5)$$

Given the noise variance, the channel estimation MSE is solely determined by the value of $\text{Tri}(\mathbf{X}_t)$. In order to minimize the value of $\text{Tri}(\mathbf{X}_t)$, the training sequences transmitted from

²The number of Rx antennas has no effects on the training requirement, and is therefore removed from the expression of the MSE.

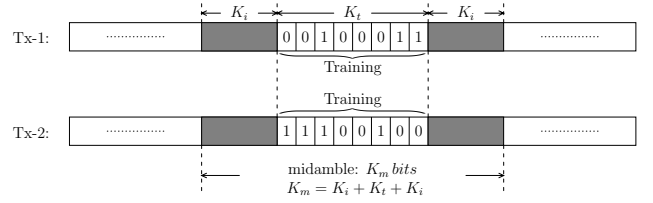


Fig. 2. Extending training with unknown data symbols

the multiple antennas must have impulse-like auto-correlation and zero cross-correlation³ [5]. For sequences fulfilling this condition, the value of $\text{Tri}(\mathbf{X}_t)$ can be written in an analytical expression⁴:

$$\text{Tri}(\mathbf{X}_t) = \frac{N_T(L+1)}{K_t - L}, \quad (6)$$

where K_t denotes the number of training symbols per Tx antenna.

Equation (6) shows two facts. First, given the system dimension, a longer training will bring a better channel estimation quality in the sense of the mean-squared error. Second, in order to guarantee a certain channel estimation quality, the length of training must increase as the number of Tx antennas increases. For example, given $L = 2$ and the desired CE quality as

$$MSE \leq 0.25 \cdot \sigma_n^2, \quad (7)$$

then the required training length will be

$$K_t \geq 26 \quad \text{for } N_T = 2, \quad (8a)$$

$$K_t \geq 98 \quad \text{for } N_T = 8, \quad (8b)$$

where the latter case can really cause a problem for a system deployed in a fast-fading environment.

B. Semi-Blind Channel Estimation

For purely training-based algorithms, a reliable estimation of a frequency-selective MIMO channel will demand a large amount of training symbols, which will considerably lower the system bandwidth efficiency. However, if we take the channel information carried by unknown data symbols into account, this situation might be changed. In the following, we introduce a semi-blind channel estimation scheme for MIMO systems, which uses training symbols together with neighboring unknown data symbols to accomplish the task of channel estimation. Compared to purely training-based schemes, the number of training symbols can be saved a lot, while the channel estimation quality is still conserved.

From now on, we consider a BPSK system with $N_T = N_R = 2$ and $L = 2$ as an illustrative example. As shown in Fig. 2⁵, we use $K_m = K_i + K_t + K_i$ symbols per Tx antenna to perform JLSCE instead of using just K_t training symbols. Let \mathbf{X}_t denote the matrix formed by training symbols, and \mathbf{X}_m denote the matrix formed by training and neighboring

³Given a finite alphabet constellation, a certain training length and a certain system dimension, sequences fulfilling this condition may not exist.

⁴BPSK modulation is assumed throughout this paper for simplicity.

⁵Within this paper, all data structures are shown in the form of binary sequences before BPSK modulation.

data symbols. Exhaustive searches show that for sure we will have the following relationship (cf. Tab. D):

$$\text{Tri}(\mathbf{X}_m) < \text{Tri}(\mathbf{X}_t), \quad \text{if } K_i > 0, \quad (9)$$

independent of the values of unknown data symbols. Furthermore, if we include more and more unknown data symbols into \mathbf{X}_m , the average value of $\text{Tri}(\mathbf{X}_m)$ becomes smaller and smaller (cf. Fig. 5):

$$\text{mean}\{\text{Tri}(\mathbf{X}'_m)\} < \text{mean}\{\text{Tri}(\mathbf{X}_m)\}, \quad \text{if } K'_i > K_i. \quad (10)$$

Above two equations show that the channel estimation quality can be improved by extending the ‘training’ sequence with data symbols, given that these data symbols are all known at the receiver.

However, the data symbols are actually unknown at receiver. Before we can get a better channel estimation, we have to estimate the data symbols correctly. But the problem is that, in order to estimate data symbols correctly, we need a perfect channel knowledge in the first place, which sounds like a self-contradiction. To solve this problem, we adopt the technique of turbo processing:

1. $i = 0: \hat{\mathbf{H}}_0 = \mathbf{Y} \cdot \mathbf{X}_t^\dagger$
2. $i = i + 1$
 - * $\hat{\mathbf{X}}_{m,i} = \text{JVD}(\mathbf{Y}_m | \hat{\mathbf{H}}_{i-1})$
 - * $\hat{\mathbf{H}}_i = \mathbf{Y}_m \cdot \hat{\mathbf{X}}_{m,i}^\dagger$
3. Repeat 2 until $(\hat{\mathbf{H}}_i, \hat{\mathbf{X}}_{m,i}) = (\hat{\mathbf{H}}_{i-1}, \hat{\mathbf{X}}_{m,i-1})$,

where i is the iteration index, and $\text{JVD}(\mathbf{Y} | \hat{\mathbf{H}}_{i-1})$ denotes data estimation by means of the *joint Viterbi detector* (JVD) [6][7] given the channel knowledge $\hat{\mathbf{H}}_{i-1}$. At the beginning, a coarse channel estimation is done by performing JLSCE over the short set of training sequences. Then, this channel knowledge is used to get an estimate of neighboring data symbols via the JVD, during which training symbols serve as a priori information. With the estimated data symbols, now we get a virtual training which is longer than the original true training sequence. Performing JLSCE over this virtual training, a new channel estimate (hopefully with lower MSE) is obtained. By repeating this procedure again and again, the channel estimation quality can be improved step by step. This statement is supported by the simulation results provided in Sec. V.

A similar iterative joint data/channel estimation algorithm has been proposed by S. Talwar in [8] for the purpose of blind equalization, which is named *iterative least-squares with projection* (ILSP). In ILSP, both data and channel estimation are done by performing an orthogonal projection in a linear space. Due to the full-rank requirement of the channel matrix, ILSP can only be applied in the situation where $N_R \geq N_T(L + 1)$, which is difficult to meet when the channel is frequency-selective. Furthermore, the small dimension of the channel matrix will produce significant noise enhancement in the procedure of data estimation, and consequently degrade the convergence property of the algorithm. In contrast, by using nonlinear techniques such as the Viterbi algorithm for data estimation, the above mentioned two problems are both removed. In accordance with the context in [8], we term the newly proposed algorithm as *iterative least-squares with Viterbi* (ILSV) algorithm.

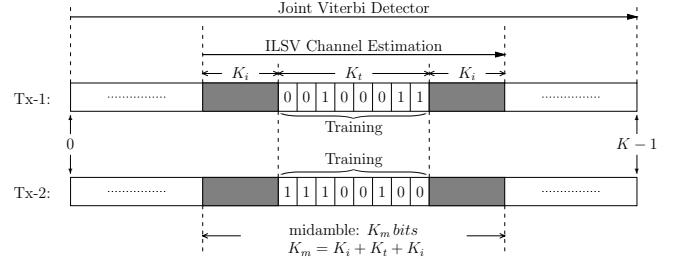


Fig. 3. JVD with SBCE

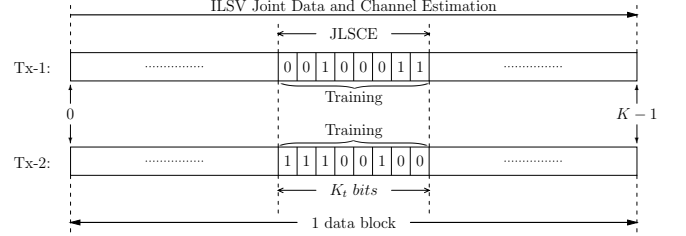


Fig. 4. Joint sequence/channel estimation by ILSV algorithm

IV. PROPOSED RECEIVER STRUCTURE

With the semi-blind channel estimation algorithm, we propose two receiver structures. For the first one, channel estimation is done by the ILSV algorithm over K_m midamble symbols. Given this channel knowledge, sequence estimation of the whole data block is done by a suitable algorithm such as the JVD. For the second one, we directly use the ILSV algorithm to jointly estimate the channel coefficients and the whole block of data symbols, which means that we choose $K_m = K$. Accordingly, we name the first receiver structure ILSV-JVD (cf. Fig. 3), and the second one ILSV (cf. Fig. 4). Comparing these two structures, ILSV-JVD is of practical interest for fast-fading channels, especially when the channel coherence interval is much shorter than the block length. While given block-fading channels, ILSV is more suitable than ILSV-JVD and will provide better performance at the price of higher computational complexity.

V. NUMERICAL RESULTS

We choose a BPSK system with two transmit antennas and two receive antennas as a test platform, and assume that the effective memory length of all sub-channels equals to two. The assumption of block-fading is also given.

A. Trace

Given a set of training sequences, we extend them by adding random data symbols at both sides, as showed in Fig. 2. As the appended data symbols are totally random, one may doubt that if the newly formed virtual training could be even worse than the original one. To give an answer, we provide two simulation results. First, given a certain training, the mean values of $\text{Tri}(\mathbf{X}_m)$ under different choice of K_m are plotted in Fig. 5. This result tells us that the newly formed virtual training matrix \mathbf{X}_m is better than \mathbf{X}_t on average. Second, given a certain training and a certain value of K_i , we check $\text{Tri}(\mathbf{X}_m)$ under all possible value combinations of the random data symbols, and measure its maximum, minimum and mean

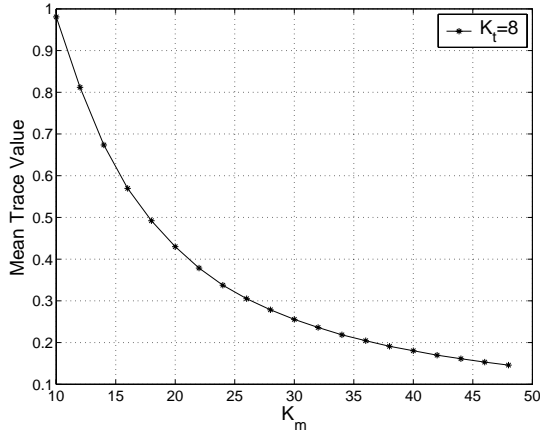


Fig. 5. Mean value of $\text{Tri}(\mathbf{X}_m)$

K_i	0	1	2	3	4
$K_m = K_t + 2 \cdot K_i$	8	10	12	14	16
$\min \{\text{Tri}(\mathbf{X}_m)\}$	1.2000	0.9333	0.6667	0.5417	0.4444
$\max \{\text{Tri}(\mathbf{X}_m)\}$	1.2000	1.0211	0.9939	0.9830	0.9770
$\text{mean} \{\text{Tri}(\mathbf{X}_m)\}$	1.2000	0.9799	0.8079	0.6737	0.5702

TABLE I
STATISTICS OF $\text{Tri}(\mathbf{X}_m)$

value, which are listed in Tab. I. A nice result is that even the maximum value of $\text{Tri}(\mathbf{X}_m)$ is smaller than $\text{Tri}(\mathbf{X}_t)$, for all $K_i > 0$. Furthermore, the average, maximum and minimum value of $\text{Tri}(\mathbf{X}_m)$ all decrease monotonously as K_i increases. We conclude that the CE performance can be certainly improved by using virtual training, as long as the included data symbols are correctly estimated at the receiver.

B. Channel Estimation

The performances of training-based and semi-blind channel estimation algorithms are depicted in Fig. 6. As we can see, for purely training-based JLSCE, a training length reduction from 26 symbols/Tx to 8 symbols/Tx leads to a 7 dB SNR penalty w.r.t. MSE. Exploiting the channel information carried by data symbols, this penalty is eliminated or at least mitigated by the ILSV algorithm. The more the data symbols we take into account, the higher the performance improvement is achieved. Particularly, when we choose $K_t = 8$ and $K_m = 32$, the CE quality of the ILSV algorithm is already comparable to the one of training-based JLSCE with $K_t = 26$ within a wide range of SNR. Due to the random property of data symbols, in order to achieve the same CE performance, a virtual training must be longer than an optimal training, but hopefully not much longer. The nonlinear behavior of the MSE curves of the ILSV algorithm in the low SNR range comes from the errors in data symbol estimation.

Another result is given in Fig. 7. Now the value of K_m is fixed to be 32, while the iteration number of the ILSV is chosen to be different values. An interesting phenomenon is that 99% of MSE reduction comes from the first iteration. This fact tells us that, if the value of K_m is not very large, we can rather fix the iteration number of the ILSV algorithm to be "1" in order to reduce the computational load.

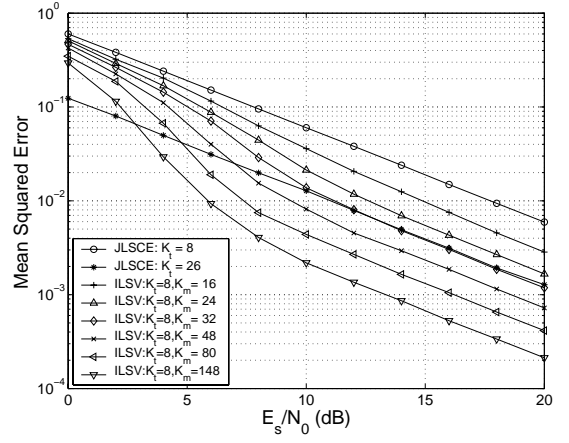


Fig. 6. MSE vs. E_s/N_0 for ILSV algorithm

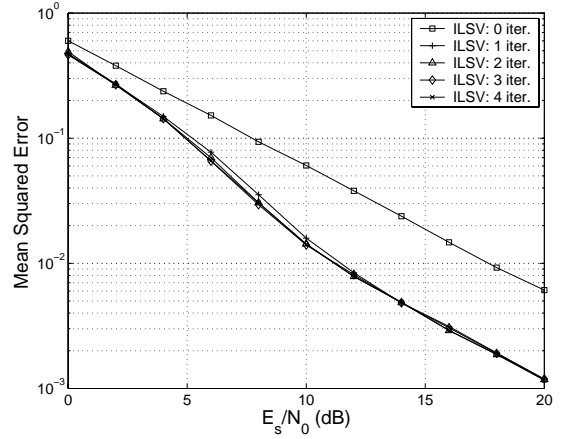


Fig. 7. MSE vs. E_s/N_0 for ILSV with different number of iterations, $K_t = 8$, $K_m = 32$

C. Channel Equalization

Joint Viterbi detector is optimal for MIMO channel equalization [6][7] and practically feasible when $N_T(L+1)$ has a moderate value. Therefore, we choose JVD as the sequence estimator for our 2×2 test system. In order to give a fair performance comparison between systems with different training lengths, the energy consumed by training symbols is subtracted from the energy of the data symbols. That is, the total available energy is fixed for a certain amount of data symbols. Thereby, a larger number of training symbols will lead to a lower SNR per symbol (E_s/N_0).

Shown by our simulation results, the MSE quality improvement provided by the semi-blind algorithm also brings a considerable BER reduction at the output of the equalizer. In Fig. 8, the ILSV-JVD receiver with $K_t = 8$ shows an approximately 2 dB SNR gain compared to the JLSCE-JVD receiver with $K_t = 8$. The only difference between these two equalization schemes is that one uses a semi-blind CE algorithm and the other uses a purely training-based CE algorithm. Given a block-fading channel, the channel coherence interval equals to the length of the data block, therefore we can actually choose $K_m = K$. According to the results in Fig. 6, such a large value of K_m should bring a high-quality channel estimate and further improve the system performance.

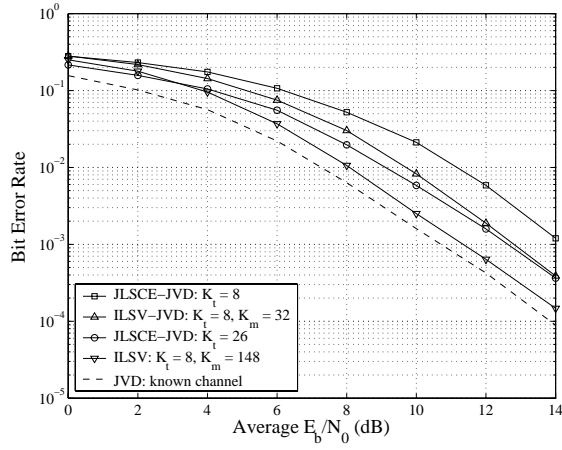


Fig. 8. BER vs. E_b/N_0 , $K = 148$, $N_T = N_R = 2$, and $L = 2$.

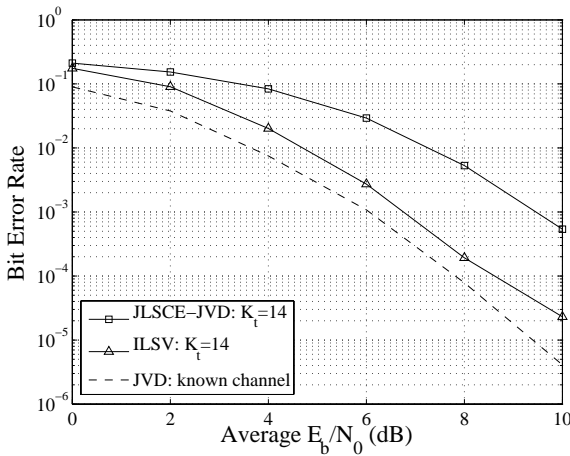


Fig. 9. BER vs. E_b/N_0 , $K = 148$, $N_T = N_R = 4$, and $L = 2$.

This is attested by the curves of the ILSV⁶ receiver in Fig. 8. When $K = 148$, the performance of the ILSV receiver is better than the JLSCE-JVD receiver with $K_t = 26$ within a wide range of SNR, and is even close to the curve of the JVD receiver with known channel coefficients. This result proves that by employing semi-blind channel estimation algorithms, the length of the training sequences required for desirable BER performance can be reduced significantly.

To further verify the effectiveness of the proposed algorithm, the simulation results for a system with $N_T = N_R = 4$ and $L = 2$ is provided in Fig. 9, and the simulation results for a system with $N_T = N_R = 8$ and $L = 0$ is also provided in Fig. 10. These results show that the semi-blind channel estimation algorithm works both for frequency-selective and flat-fading MIMO systems with various number of antennas.

VI. CONCLUSIONS

For conventional approaches, very long training will be needed in order to obtain a good estimate of a frequency-selective MIMO channel. In this paper, we investigated the possibility of reliable channel estimation when only short

⁶As the ILSV algorithm actually estimates the data and the channel jointly, the task of channel equalization is already done by it when $K_m = K$ is chosen, therefore we name the receiver structure as ILSV under this situation.

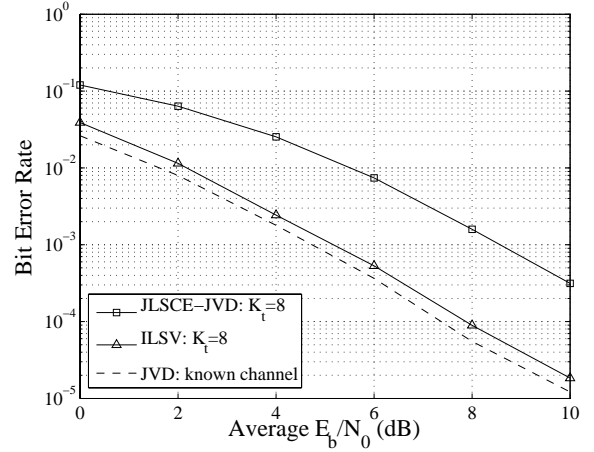


Fig. 10. BER vs. E_b/N_0 , $K = 148$, $N_T = N_R = 8$, and $L = 0$.

training is available. The simulation results show that, by utilizing iterative joint data/channel estimation, the requirements on the training length can be largely relaxed. Generally speaking, given the semi-blind channel estimation algorithm, the demanded training length is now only $N_T(L + 1) + L$, which is affordable even when N_T is rather large.

Within this paper, we used JLSCE and JVD as the channel estimation algorithm and the data estimation algorithm, respectively. As a matter of fact, any suitable data/channel estimation algorithm can be utilized to improve the channel estimation quality in the way described in Sec. III-B. The computational complexity of JLSCE is proportional to $N_T^3(L + 1)^3$, and should be manageable in most cases. However, the computational complexity of JVD is exponential in $N_T(L + 1)$, which will obviously bring problems when the system has many Tx antennas. Therefore, data detection algorithms with reduced complexity need to be investigated. Sequential detection techniques which combine linear and non-linear equalization algorithms should be a potential solution. Furthermore, the redundancy introduced by channel coding might also be utilized in the procedure of data estimation. These topics are subject to further research.

REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585-595, Nov. 1999.
- [2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41-59, 1996.
- [3] S.N. Crozier, D.D. Falconer, and S.A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEEE Proc.-F*, vol. 138, no. 4, pp. 371-378, Aug. 1992.
- [4] P. Ranta, A. Hottinen, and Z.-C. Honkasalo, "Co-channel interference cancellation receiver for TDMA mobile systems," in *Proc. IEEE ICC'95*, Seattle, pp. 17-21, Jun. 1995.
- [5] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 384-391, Mar. 2003.
- [6] K. Giridhar, S. Chari, J. Shynk, R. Gooch, and D. Artman, "Joint estimation algorithms for cochannel signal demodulation," in *Proc. IEEE ICC'93*, Geneva, pp. 1497-1501, May 1993.
- [7] K. Giridhar, J. Shynk, A. Mathur, S. Chari, and R. Gooch, "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Trans. Comm.*, vol. 45, no. 4, pp. 473-484, Apr. 1997.
- [8] S. Talwar, *Blind Space-Time Algorithms for Wireless Communication Systems*, Ph.D. Thesis, Stanford University, Jan. 1996.