Performance Analysis of Maximum-Likelihood Semiblind Estimation of MIMO Channels

Tianbin Wo and Peter Adam Hoeher

Faculty of Engineering
University of Kiel
Kaiserstr. 2, D-24143 Kiel, Germany
Email: {wtb,ph}@tf.uni-kiel.de

Ansgar Scherb and Karl-Dirk Kammeyer

Department of Communications Engineering
University of Bremen
Otto-Hahn-Allee, D-28359 Bremen, Germany
Email:{scherb,kammeyer}@ant.uni-bremen.de

Abstract— **Iterative channel estimation and data detection is a very useful method to improve the channel estimation quality without sacrificing the bandwidth efficiency. Since both the known training symbols (non-blind) and the unknown data symbols (blind) are used for channel estimation, corresponding techniques are referred to as semiblind. If the channel estimator and data detector are both optimal in the sense of maximum-likelihood criterion, we may call the algorithm as *maximum-likelihood* (ML) *semiblind channel estimation* (SBCE). This paper deals with ML-SBCE for frequency-flat multi-input multi-output systems with focus on the channel estimation *mean squared error* (MSE) analysis. Through semi-analytical efforts, we will show that ML-SBCE is biased at low SNR and tends to be unbiased at high SNR. The reasons of biasing are the erroneous data detection and the correlation between the noise and the detection errors. Besides, we will show that the MSE performance of ML-SBCE is also influenced by the noise-error correlation. Based on these analysis, possibilities to compensate the biasing as well as improve the MSE performance are pointed out.**

I. INTRODUCTION

The dramatic capacity gain of *multi-input multi-output* (MIMO) channels w.r.t. *single-input single-output* (SISO) channels have been shown by numerous literatures [1]-[3], in an environment rich of multipath propagation. However, in practice it is a challenging task to fully exploit the MIMO channel capacity. One difficulty exists in the contradictory requirements of reliable channel estimation and high system bandwidth efficiency. Due to multi-antenna interferences, a MIMO receiver is generally more sensitive to the channel estimation errors than a SISO one, meanwhile the number of channel coefficients to be estimated is also much larger than that of a SISO system. Therefore, in order to achieve a desirable *bit error ratio* (BER) performance a long training will be necessary. On the other hand, the demand of high bandwidth efficiency forces the training to be as short as possible. Hassibi *et al.* showed in [4] that pure training-based channel estimation can be highly suboptimal from the information theoretic point of view. In comparison, semiblind channel estimation tries to extract the channel state information carried by all observations, and is able to achieve very low MSE with using just a few training symbols. Exchanging the information between the channel estimator and the data detector iteratively,

a SBCE algorithm refines the quality of channel estimates and data decisions in a recursive manner. Simulation results of such channel estimators can be found in a number of literatures [5]-[8], where the huge SNR gain by using SBCE was shown. While, theoretical analysis available from the literature so far is still not sufficient. Buzzi *et al.* discussed in [9] the behaviour of SBCE algorithms with linear data detectors, and provided closed form of formulas for predicting the MSE performance. However, one important assumption made by the authors, that the correlation between the noise samples and the detection errors is negligible, is actually not appropriate. Indeed, our investigation shows that this noise-error correlation has a strong impact both on the biasing and the MSE of semiblind channel estimators.

In this paper we provide analytical results for the SBCE algorithm with a ML channel estimator and a ML data detector. As ML detection provides the performance limit for all data detectors, the results of ML-SBCE can be taken as the upper bound for algorithms with other types of detectors. Since a literal derivation of the mean value and the MSE is mathematically far too complicated, many approximations are made in order to get neat expressions. Nevertheless, computer simulations will show that these approximations work very well and are accurate for the SNR range of practical interest.

The paper is organized as follows: Sec. II gives the channel model, and Sec. III introduces the ML-SBCE algorithm. In Sec. IV, detailed performance analysis together with numerical results are provided, and Sec. V concludes the paper.

II. CHANNEL MODEL

Let N_R denote the number of receive (Rx) antennas and N_T the number of transmit (Tx) antennas, the equivalent discrete-time model of a $(N_R \times N_T)$-MIMO channel (including transmit and receive filter, physical channel and baud-rate sampling) is given by

$$\mathbf{r}(k) = \mathbf{H}(k)\mathbf{s}(k) + \mathbf{n}(k), \tag{1}$$

where k is the discrete time index. The channel input vector $\mathbf{s}(k) \in \mathbb{C}^{N_T \times 1}$ consists of BPSK[1] modulated symbols, and

---

[1]BPSK modulation is assumed throughout the paper for the sake of simple analytical expressions.

$\mathbf{r}(k) \in \mathbb{C}^{N_R \times 1}$ denotes the channel output. $\mathbf{n}(k) \in \mathbb{C}^{N_R \times 1}$ is a zero-mean white Gaussian noise vector with covariance $\mathrm{E}\{\mathbf{n}(k)\mathbf{n}^H(k)\} = \sigma_n^2 \mathbf{I}_{N_R}$, where $\mathbf{I}_{N_R}$ is the identity matrix of order N_R. The channel matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is assumed to be constant over K symbol periods (block fading), so that the transmission of K consecutive vector-symbols can be compactly written as

$$\mathbf{R} = \mathbf{HS} + \mathbf{N}, \tag{2}$$

where $\mathbf{R} = [\mathbf{r}(0), \cdots, \mathbf{r}(K-1)]$, $\mathbf{S} = [\mathbf{s}(0), \cdots, \mathbf{s}(K-1)]$ and $\mathbf{N} = [\mathbf{n}(0), \cdots, \mathbf{n}(K-1)]$.

Assuming that the training is inserted into the preamble of each data burst, the symbol matrix can be expressed as

$$\mathbf{S} = [\mathbf{S}_T, \mathbf{S}_I], \tag{3}$$

where $\mathbf{S}_T$ contains training symbols spanning K_T time slots, and $\mathbf{S}_I$ contains info symbols spanning K_I time slots. Correspondingly, the channel output can be also written as

$$\mathbf{R} = [\mathbf{R}_T, \mathbf{R}_I] = [\mathbf{HS}_T + \mathbf{N}_T, \mathbf{HS}_I + \mathbf{N}_I]. \tag{4}$$

We choose this notation for easy algorithm description and performance analysis in the remainder of this paper.

III. SEMIBLIND CHANNEL ESTIMATION

Traditional algorithms use only the training to perform channel estimation, while a semiblind channel estimator takes the data symbols also into account. Since the data symbols are practically unknown, before they can be used for channel estimation, the receiver has to perform a detection in advance. Thus, the task of channel estimation changes into joint estimation of channel and data symbols. Borrowing ideas from Turbo processing, this procedure can be done in an iterative manner:

1) Initial training-based channel estimation;
2) Given the channel knowledge, perform data detection;
3) Given the data knowledge, perform channel estimation by taking the whole data block as a virtual training;
4) Repeat step 2 and step 3 until a certain stopping criterion is reached.

By refining the channel estimate and the data decisions in a recursive manner, considerable performance gain can be achieved step by step. In the following, the issues of channel estimation and data detection will be tackled respectively.

A. Channel Estimation

The *least-squares* (LS) algorithm and the *minimum mean squared error* (MMSE) algorithm are quite popular for channel estimation. Given perfect knowledge of the transmitted symbols (e.g. the training), a LS algorithm delivers unbiased channel estimates, while a MMSE algorithm delivers biased ones. For the sake of simplicity, in this paper we restrict ourself to the LS algorithm on account of its unbiasedness. A LS channel estimator minimizes the distance between the channel output and its noiseless hypothesis given by

$$\widehat{\mathbf{H}} = \arg\min_{\widetilde{\mathbf{H}}} \left\{ \|\mathbf{R} - \widetilde{\mathbf{H}}\mathbf{S}\|_F^2 \right\} = \mathbf{RS}^H \left(\mathbf{SS}^H\right)^{-1}, \tag{5}$$

which in turn maximizes the likelihood function $p(\mathbf{R}|\widetilde{\mathbf{H}}, \mathbf{S})$, and thus is often called a ML channel estimator.

For SBCE, the initial channel estimation is performed over the training symbols only:

$$\widehat{\mathbf{H}}_{\text{init}} = \mathbf{R}_T \mathbf{S}_T^H \left(\mathbf{S}_T \mathbf{S}_T^H\right)^{-1}, \tag{6}$$

while in the later iteration, the knowledge of data symbols will also be utilized:

$$\widehat{\mathbf{H}} = \mathbf{R}\widehat{\mathbf{S}}^H \left(\widehat{\mathbf{S}}\widehat{\mathbf{S}}^H\right)^{-1}, \tag{7}$$

with $\widehat{\mathbf{S}} = [\mathbf{S}_T, \widehat{\mathbf{S}}_I]$. If the amount of detection errors in $\widehat{\mathbf{S}}_I$ is small enough, (7) will hopefully produce a better channel estimate than (6) in the sense of lower MSE.

B. Data Detection

The classical nonlinear receiver is the ML detector which searches for the most likely data sequence according to

$$\widehat{\mathbf{S}}_{\text{ML}} = \arg\min_{\widetilde{\mathbf{S}} \in \mathcal{S}} \left\{ \|\mathbf{R} - \widehat{\mathbf{H}}\widetilde{\mathbf{S}}\|_F^2 \right\}, \tag{8}$$

where $\mathcal{S}$ denotes the set of realizations with $|\mathcal{S}| = 2^{N_T \times K}$. The ML detector delivers perfect performance compared to linear algorithms, such as zero-forcing and MMSE, but has a complexity proportional to 2^{N_T}. For real-world applications, the most attractive solutions fall into a blended class, that can combine the advantages of linear and nonlinear algorithms, and achieve desirable performance with affordable complexity. Due to limited space, we restrict ourself to the SBCE algorithm with a LS channel estimator and a ML data detector.

IV. PERFORMANCE ANALYSIS OF CHANNEL ESTIMATION

In this section, we provide numerical results and detailed analysis concerning the biasing and the MSE of the maximum-likelihood semiblind channel estimator.

A. On the Biasing of ML-SBCE

It is easy to find that equation (6) always delivers an unbiased channel estimate:

$$\mathrm{E}\{\widehat{\mathbf{H}}_{\text{init}}\} = \mathbf{H} + \mathrm{E}\{\mathbf{N}_T\}\mathbf{S}_T^H \left(\mathbf{S}_T \mathbf{S}_T^H\right)^{-1} = \mathbf{H}. \tag{9}$$

While, it remains to be an interesting question whether $\widehat{\mathbf{H}}$ in (7) is biased or not. Given sufficient large block length, the following approximation

$$\widehat{\mathbf{S}}\widehat{\mathbf{S}}^H \approx K\mathbf{I}_{N_T} \tag{10}$$

should be valid, then $\widehat{\mathbf{H}}$ can be rewritten as

$$\widehat{\mathbf{H}} \approx \frac{1}{K}\mathbf{R}\widehat{\mathbf{S}}^H = \frac{1}{K}\left(\mathbf{HS}\widehat{\mathbf{S}}^H + \mathbf{N}\widehat{\mathbf{S}}^H\right). \tag{11}$$

Given optimal training (cf. [10]), and let P_s denote the symbol error ratio at the output of the data detector[2], we have

$$\mathbf{S}_T \mathbf{S}_T^H = K_T \mathbf{I}_{N_T} \tag{12}$$

$$\mathrm{E}\{\mathbf{S}_I \widehat{\mathbf{S}}_I^H\} = (1 - 2P_s)K_I \mathbf{I}_{N_T}, \tag{13}$$

---

[2]Here we assume that the detection error ratios for the data from each Tx antenna are equivalent, which is true on average.
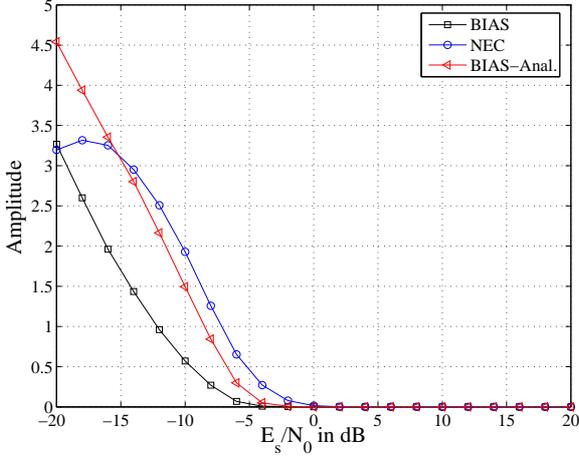
Fig. 1. BIAS vs. SNR, with $N_\text{T} = N_\text{R} = 4$, $K = 200$, $K_\text{T} = 4$.

and consequently

$$\begin{aligned}
\text{E}\{\mathbf{S}\widehat{\mathbf{S}}^H\} &= \text{E}\{\mathbf{S}_\text{T}\mathbf{S}_\text{T}^H\} + \text{E}\{\mathbf{S}_\text{I}\widehat{\mathbf{S}}_\text{I}^H\} \\
&= \big(K_\text{T} + (1 - 2P_s)K_\text{I}\big) \cdot \mathbf{I}_{N_\text{T}}.
\end{aligned} \tag{14}$$

Noting that the noise samples are independent with the training and the data symbols, and let $\mathbf{E}_\text{I} = \widehat{\mathbf{S}}_\text{I} - \mathbf{S}_\text{I}$ denote the symbol estimation error, we obtain

$$\begin{aligned}
\text{E}\{\mathbf{N}\widehat{\mathbf{S}}^H\} &= \text{E}\{\mathbf{N}_\text{T}\mathbf{S}_\text{T}^H + \mathbf{N}_\text{I}(\mathbf{S}_\text{I} + \mathbf{E}_\text{I})^H\} \\
&= \text{E}\{\mathbf{N}_\text{I}\mathbf{E}_\text{I}^H\}.
\end{aligned} \tag{15}$$

Combining (14), (15) and (11), the mean value of $\widehat{\mathbf{H}}$ can be finally approximated by

$$\text{E}\{\widehat{\mathbf{H}}\} \approx \frac{K_\text{T} + (1 - 2P_s)K_\text{I}}{K}\mathbf{H} + \frac{1}{K}\text{E}\{\mathbf{N}_\text{I}\mathbf{E}_\text{I}^H\}, \tag{16}$$

where the right term actually indicates the correlation between the noise samples and the detection errors. It is obvious that $\widehat{\mathbf{H}}$ tends to be unbiased for high SNR as both P_s and $\mathbf{E}_\text{I}$ become zero, while biased at low SNR. Contrary to the statements in [9], the noise-error correlation $\text{E}\{\mathbf{N}_\text{I}\mathbf{E}_\text{I}^H\}$ should not be assumed to be zero, since the noise is exactly the cause of detection errors. This statement will be attested by the following numerical results.

Given a randomly chosen channel matrix, Fig. 1 shows the simulation results of the degree of biasing defined as

$$\textbf{BIAS} \doteq \big\|\mathbf{H} - \text{E}\{\widehat{\mathbf{H}}\}\big\|_\text{F}^2 \tag{17}$$

and the degree of noise-error correlation defined as

$$\textbf{NEC} \doteq \big\|\frac{1}{K}\text{E}\{\mathbf{N}_\text{I}\mathbf{E}_\text{I}^H\}\big\|_\text{F}^2 \tag{18}$$

versus SNR, respectively. "BIAS-Anal." in Fig. 1 denotes the analytical value of BIAS by replacing $\text{E}\{\widehat{\mathbf{H}}\}$ in (17) with the approximation given in (16). As expected, ML-SBCE is biased at low SNR while unbiased at high SNR, and the degree of biasing decreases as the noise power decreases. The noise-error correlation is not negligible, and indeed its value

is significant w.r.t. BIAS. An interesting observation is that the biasing due to NEC is partially compensated by the left term in (16), which means that the correlation pattern between the noise and the errors is determined by the channel matrix. The distance between BIAS and "BIAS-Anal." arises from the approximation given in (10).

B. Biasing Compensation

Since certain types of data detectors may benefit from unbiased channel estimates, it makes sense to compensate the biasing of a semiblind channel estimator. As we have seen from equation (16), the ML-SBCE is biased by a positive scalar and an additive noise-error correlation. Given a system with low-rate channel code, if we would feedback the decoder output to the channel estimator, then the noise samples and detection errors will be quasi independent. Under such a situation, the right term in (16) will disappear, and we can easily get an unbiased semiblind channel estimator given by

$$\widehat{\mathbf{H}} \triangleq \frac{K}{K_\text{T} + (1 - 2P_s)K_\text{I}}\mathbf{R}\widehat{\mathbf{S}}^H(\widehat{\mathbf{S}}\widehat{\mathbf{S}}^H)^{-1}. \tag{19}$$

Due to limited space, we would refer interested readers to [11] for further details on this topic.

C. Mean Squared Error

For MIMO systems, the channel estimation MSE can be defined as[3]

$$\text{MSE} \doteq \text{E}\big\{\|\widehat{\mathbf{H}} - \mathbf{H}\|_\text{F}^2\big\}. \tag{20}$$

Before we start to assess the performance of SBCE, we may intuitively set a lower bound and an upper limit for the MSE. First, the lower bound should be given by the LS estimator when perfect data knowledge is available, which is denoted by *data-based channel estimation* (DBCE) hereafter. Second, if the MSE performance of SBCE is even worse than that of the pure *training-based channel estimation* (TBCE), then it is meanless to perform SBCE. In fact, the MSE curves of TBCE and DBCE give the *Cramer-Rao lower bound* (CRLB) for any unbiased semiblind channel estimators at low and high SNR (cf. [12]).

Fig. 2 demonstrates the performance of ML-SBCE given different number of receive antennas. Without using more training symbols, the ML-SBCE algorithm achieves a huge SNR gain w.r.t. the pure training-based one. As shown in Fig. 2, ML-SBCE works very well even when $N_\text{R} < N_\text{T}$, and is able to provide better channel estimates as N_R increases. It should be mentioned that a TBCE algorithm does not benefit from diversity reception at all. As we may notice, the MSE curves of ML-SBCE touch the one of DBCE at high SNR range, and the position of this touching point shifts to the left side as N_R increases. This phenomenon may be interpreted as follows: the more the receive antennas, the smaller the SNR value that ML-SBCE needs in order to achieve the same MSE performance as if all data symbols are perfectly known

---

[3]In this paper, all MSE simulation results have been normalized by the factor of $N_\text{R} \times N_\text{T}$ for the purpose of fair comparison between systems with different number of antennas.
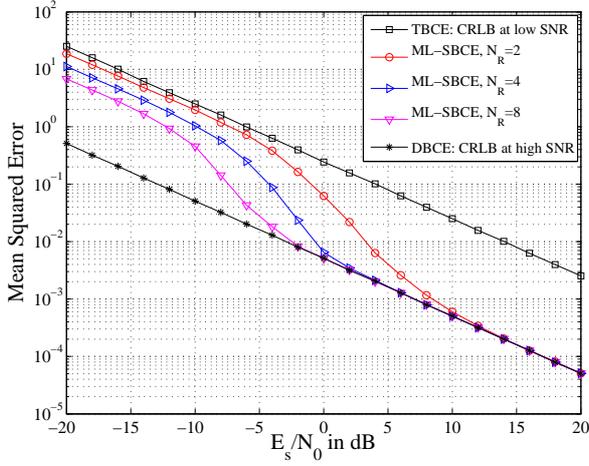
Fig. 2. MSE vs. SNR, $N_\mathrm{T} = 4$, $K = 200$, $K_\mathrm{T} = 4$.

Fig. 3. MSE analysis, with $N_\mathrm{T} = N_\mathrm{R} = 4$, $K = 200$, $K_\mathrm{T} = 4$.

at the receiver. Given $N_\mathrm{R} = N_\mathrm{T} = 4$, this touching point[4] is approximately at $SNR = 3$ dB, which deserves to be an amazing performance. Please note that the number of training symbols used here is only 4 per burst per Tx antenna, which is actually the minimum amount of training in order to perform LS channel estimation. Another important observation is that the performance of ML-SBCE may exceed the CRLB at low SNR due to biasing.

Now we try to give an analytical expression of the MSE by using the following approximation

$$\mathbf{S}_\mathrm{I}\widehat{\mathbf{S}}_\mathrm{I}^H \approx (1 - 2P_s)K_\mathrm{I}\mathbf{I}_{N_\mathrm{T}}, \qquad (21)$$

which is accurate if $N_\mathrm{T} = 1$. After some derivations (cf. Appendix A), we will finally obtain

$$\mathrm{MSE} \approx \underbrace{\left(\frac{2P_sK_\mathrm{I}}{K}\right)^2 \|\mathbf{H}\|_\mathrm{F}^2}_{\alpha}$$
$$- \underbrace{\frac{4P_sK_\mathrm{I}}{K^2}\mathrm{tr}\Big\{\mathrm{E}\big\{\mathbf{N}_\mathrm{I}\mathbf{E}_\mathrm{I}^H\big\}\mathbf{H}^H\Big\}}_{\beta} \qquad (22)$$
$$+ \underbrace{\frac{N_\mathrm{R}N_\mathrm{T}}{K}\sigma_n^2}_{\gamma} + \underbrace{\frac{1}{K^2}\mathrm{E}\big\{\|\mathbf{N}_\mathrm{I}\mathbf{E}_\mathrm{I}^H\|_\mathrm{F}^2\big\}}_{\delta}.$$

Let α, β, γ and δ represent the corresponding items in (22) respectively, Fig. 3 plots their values versus E_S/N_0 and compares them to the true MSE values. Surprisingly, α equals to β for all SNR values, and this insight provides us the following empirical conclusion

$$\mathrm{tr}\Big\{\Re\big\{\mathrm{E}\big\{\mathbf{N}_\mathrm{I}\mathbf{E}_\mathrm{I}^H\big\}\mathbf{H}^H\big\}\Big\} \approx P_sK_\mathrm{I}\mathrm{tr}\Big\{\mathbf{H}\mathbf{H}^H\Big\}, \qquad (23)$$

which again tells us that the pattern of noise-error correlation is solely determined by the channel matrix, when a ML detector

[4]The mathematical expression for predicting the position of this touching point has recently been found, while due to limited space, we would refer interested readers to [13].
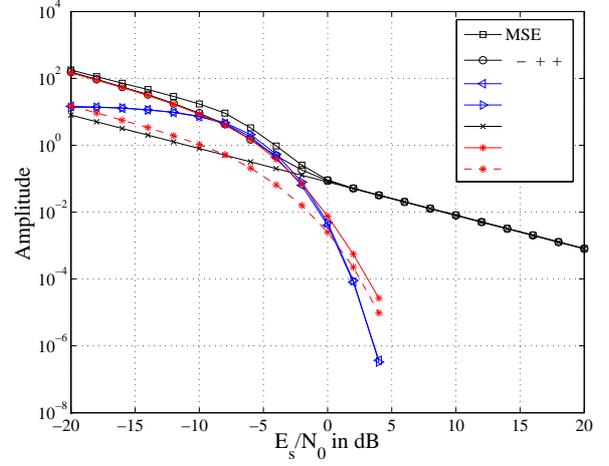
is used. Now, the MSE expression can be further simplified as

$$\mathrm{MSE} \approx \frac{N_\mathrm{R}N_\mathrm{T}}{K}\sigma_n^2 + \frac{1}{K^2}\mathrm{E}\big\{\|\mathbf{N}_\mathrm{I}\mathbf{E}_\mathrm{I}^H\|_\mathrm{F}^2\big\}, \qquad (24)$$

where the left term coincides with CRLB for high SNR [13] and gives a tight lower bound of the MSE. At high SNR, δ tends to be zero, and thereby the ML-SBCE touches the lower bound. While at very low SNR, the value of δ seems dominant and tends to be linear w.r.t. the noise power. We may also notice that, it is δ that determines the position where the MSE line at high SNR transits to the MSE line at low SNR.

Buzzi *et al.* in [9] assumed the noise-error correlation to be zero, which leads to

$$\frac{1}{K^2}\mathrm{E}\big\{\|\mathbf{N}_\mathrm{I}\mathbf{E}_\mathrm{I}^H\|_\mathrm{F}^2\big\} \overset{?}{=} \underbrace{\frac{1}{K^2}4P_sK_\mathrm{I}N_\mathrm{R}N_\mathrm{T} \cdot \sigma_n^2}_{\zeta}, \qquad (25)$$

by noting that the elements of $\mathbf{E}_\mathrm{I}$ belongs to $\{0, +2, -2\}$. The curve of ζ versus E_S/N_0 is also plotted in Fig. 3, and obviously we have $\delta > \zeta$ for the whole SNR range. This result delivers the message that the noise-error correlation is not negligible for the MSE. However, as equation (25) does not hold, a suitable analytical expression remains to be an open question up to the moment of writing this paper.

Nevertheless, the distance between δ and ζ gives us a hint: if we could make the detection errors independent with the noise samples, then a channel estimator with lower MSE might result. Hopefully, this is the true situation when the decoder output is fed back for channel estimation. For more details, we would refer readers to [11].

D. Bit Error Ratio

The ultimate goal of conducting semiblind channel estimation is to improve the system BER performance. Fig. 4 compares the BER performance of systems with and without SBCE and also the case when perfect channel knowledge is available. The improvement in channel estimation quality from executing SBCE does yield significant BER reduction.
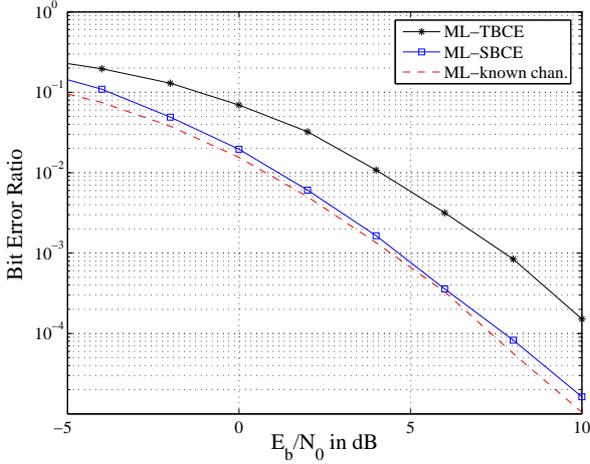
Fig. 4. BER vs. E_b/N_0, with $N_{\mathrm{R}} = N_{\mathrm{T}} = 4$, $K = 200$, $K_{\mathrm{T}} = 4$.

V. CONCLUSIONS

In this paper, the behaviour of the maximum-likelihood semiblind channel estimator is studied for MIMO systems. Analytical expressions for the biasing and the MSE are provided, together with illustrative simulation results. As we have shown, ML-SBCE is biased at low SNR while tends to be unbiased at high SNR, and both its mean value and MSE are highly dependent on the cross-correlation between the noise samples and the detection errors. Based on these analysis, possibilities to compensate the biasing and improve the MSE performance are pointed out.

APPENDIX

A. Derivation of the MSE Expression

According to (20), the MSE can be written as

$$
\begin{aligned}
\mathrm{MSE} &= \mathrm{E}\Big\{\mathrm{tr}\big\{(\mathbf{H} - \widehat{\mathbf{H}})(\mathbf{H} - \widehat{\mathbf{H}})^H\big\}\Big\} \\
&= \|\mathbf{H}\|_{\mathrm{F}}^2 - \mathrm{tr}\big\{\mathbf{H}\,\mathrm{E}\{\widehat{\mathbf{H}}^H\}\big\} \\
&\quad - \mathrm{tr}\big\{\mathrm{E}\{\widehat{\mathbf{H}}\}\mathbf{H}^H\big\} + \mathrm{E}\big\{\|\widehat{\mathbf{H}}\|_{\mathrm{F}}^2\big\}, \quad (26)
\end{aligned}
$$

where we measure the MSE w.r.t. a certain channel matrix. Given the approximation in (21) and recalling equation (11), we get the expression

$$
\widehat{\mathbf{H}} \approx \frac{K_{\mathrm{T}} + (1 - 2P_s)K_{\mathrm{I}}}{K}\mathbf{H} + \frac{1}{K}\mathbf{N}\widehat{\mathbf{S}}^H. \quad (27)
$$

For simplicity, we define

$$
\alpha = K_{\mathrm{T}} + (1 - 2P_s)K_{\mathrm{I}}, \quad (28)
$$

then the mean power of $\widehat{\mathbf{H}}$ can be approximated by

$$
\begin{aligned}
\mathrm{E}\big\{\|\widehat{\mathbf{H}}\|_{\mathrm{F}}^2\big\} &\approx \frac{1}{K^2}\mathrm{E}\big\{\|\alpha\mathbf{H} + \mathbf{N}\widehat{\mathbf{S}}^H\|_{\mathrm{F}}^2\big\} \\
&= \frac{1}{K^2}\Big\{\alpha^2\|\mathbf{H}\|_{\mathrm{F}}^2 + \alpha\,\mathrm{tr}\big\{\mathbf{H}\,\mathrm{E}\{\widehat{\mathbf{S}}\mathbf{N}^H\}\big\} \quad (29) \\
&\quad + \alpha\,\mathrm{tr}\big\{\mathrm{E}\{\mathbf{N}\widehat{\mathbf{S}}^H\}\mathbf{H}^H\big\} + \mathrm{E}\big\{\|\mathbf{N}\widehat{\mathbf{S}}^H\|_{\mathrm{F}}^2\big\}\Big\}.
\end{aligned}
$$

By inserting (27) and (29) into (26), we obtain

$$
\begin{aligned}
\mathrm{MSE} &\approx (1 - \frac{\alpha}{K})^2\|\mathbf{H}\|_{\mathrm{F}}^2 \\
&\quad + \frac{2(\alpha - K)}{K^2}\mathrm{tr}\big\{\Re\{\mathrm{E}\{\mathbf{N}\widehat{\mathbf{S}}^H\}\mathbf{H}^H\}\big\} \\
&\quad + \frac{1}{K^2}\mathrm{E}\big\{\|\mathbf{N}\widehat{\mathbf{S}}^H\|_{\mathrm{F}}^2\big\}. \quad (30)
\end{aligned}
$$

Noting that

$$
\begin{aligned}
\mathrm{E}\big\{\|\mathbf{N}\widehat{\mathbf{S}}^H\|_{\mathrm{F}}^2\big\} &= \mathrm{E}\big\{\|\mathbf{N}\mathbf{S}^H + \mathbf{N}_{\mathrm{I}}\mathbf{E}_{\mathrm{I}}^H\|_{\mathrm{F}}^2\big\} \\
&\approx \mathrm{E}\big\{\|\mathbf{N}\mathbf{S}^H\|_{\mathrm{F}}^2\big\} + \mathrm{E}\big\{\|\mathbf{N}_{\mathrm{I}}\mathbf{E}_{\mathrm{I}}^H\|_{\mathrm{F}}^2\big\} \\
&= KN_{\mathrm{R}}N_{\mathrm{T}}\sigma_n^2 + \mathrm{E}\big\{\|\mathbf{N}_{\mathrm{I}}\mathbf{E}_{\mathrm{I}}^H\|_{\mathrm{F}}^2\big\}, \quad (31)
\end{aligned}
$$

and inserting (15) and (28) into (30), we finally obtain

$$
\begin{aligned}
\mathrm{MSE} &\approx \left(\frac{2P_s K_{\mathrm{I}}}{K}\right)^2\|\mathbf{H}\|_{\mathrm{F}}^2 \\
&\quad - \frac{4P_s K_{\mathrm{I}}}{K^2}\mathrm{tr}\big\{\Re\{\mathrm{E}\{\mathbf{N}_{\mathrm{I}}\mathbf{E}_{\mathrm{I}}^H\}\mathbf{H}^H\}\big\} \\
&\quad + \frac{N_{\mathrm{R}}N_{\mathrm{T}}}{K}\sigma_n^2 + \frac{1}{K^2}\mathrm{E}\big\{\|\mathbf{N}_{\mathrm{I}}\mathbf{E}_{\mathrm{I}}^H\|_{\mathrm{F}}^2\big\}. \quad (32)
\end{aligned}
$$

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585-595, 1999.
[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311-335, 1998.
[3] B. Hochwald and T. L. Marzetta, "Capacity of a mobile mutliple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Jan. 1999.
[4] B. Hassibi, and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links? ," in *IEEE Trans. on Information Theory*, vol. 49, no. 4, pp. 951-963, April 2003.
[5] A. Medles and DTM. Slock, "Semiblind channel estimation for MIMO spatial multiplexing systems," *IEEE Vehic. Techn. Conf. (VTC)*, Fall, 2001.
[6] M. Loncar *et al.*, "Iterative joint detection, decoding, and channel estimation for dual antenna arrays in frequency selective fading," in *Proc. 5th Int. Symp. Wireless Personal Multimedia Communication*, Honolulu, HI, Oct. 2002, pp. 125-129.
[7] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, pp. 1266-1270, Aug. 2003.
[8] T. Wo and P. A. Hoeher, "Semi-blind channel estimation for frequency-selective MIMO systems," in *14-th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June, 2005.
[9] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. Vehic. Techno.*, vol. 53, no. 4, pp. 1085-1104, July 2004.
[10] 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.
[11] A. Scherb, T. Wo, P. A. Hoeher and K. D. Kammeyer, "Unbiased semiblind channel estimation for coded systems," submitted to *IEEE Vehicular Technology Conference*, Melbourne, Australia, May, 2006.
[12] A. Scherb, V. Kuehn and K. D. Kammeyer, "Cramer-Rao Lower Bound for semiblind channel estimation with respect to coded and uncoded finite-alphabet signals," in *Asilomar Conference on Signals, Systems, and Computers*, Monterey, USA, November, 2004
[13] T. Wo, A. Scherb, K. D. Kammeyer and P. A. Hoeher, "Analysis of semiblind channel estimation for FIR-MIMO systems," submitted to *The 4th International Symposium on Turbo Codes & Related Topics*, Munich, Germany, April, 2006.