PERFORMANCE OF LINEAR RECEIVERS BASED ON SUPERIMPOSED TRAINING

Abla Kammoun[1] *Karim Abed-Meraim*[1] *and Sofiène Affes*[2]

Ecole Nationale Supérieure des Télécommunications 46, rue Barrault, 75634 Paris Cedex 13, France[1]
INRS-EMT, 800, de la Gauchetière, Bureau 6900, Montreal, QC, H5A 1K6 Canada[2]
E-mails: kammoun, abed@tsi.enst.fr, affes@emt.inrs.ca

1. ABSTRACT

In this paper, we derive a performance comparison between two training-based schemes for MIMO semi-blind channel estimation. The two schemes are the conventional superimposed training scheme and the more recently proposed data-dependent superimposed pilot scheme. For both schemes, a closed-form for the outage probability and a lower bound of the bit error rate are given. We also determine for the data-dependant superimposed training scheme the optimal allocation of power between pilot and data. Once the optimal data and training power are set, we prove analytically that the optimum data-dependent superimposed scheme always outerperforms the conventional superimposed training scheme.

Key words: semi-blind channel estimation, superimposed training sequence, MIMO systems performance, linear receiver.

2. INTRODUCTION

The use of Multiple-Input Multiple-Output (MIMO) antenna systems enables high data rates without any increase in bandwidth or power consumption. However, the good performance of MIMO systems requires a priori knowledge of the channel at the receiver. In many practical systems, the receiver estimates the channel by time division multiplexing pilot symbols with the data. Although high quality of channel estimation could be achieved especially when using a large number of pilot symbols [1], this method may entail a waste of the available bandwidth. An alternative method is the conventional superimposed training. It consists in transmitting pilots and data at the same time. However, since during channel estimation, data symbols act as an input source of noise, channel estimation is affected. In [2], M. Ghogho et al propose to introduce a distorsion to the data symbols, prior to adding the known pilot. By this way, it is shown that the channel estimation performance is by far enhanced. This technique is referred to as the data-dependent superimposed training.

In this paper, we propose to derive an analytical performance comparison between the conventional and the data-dependent superimposed training schemes. We show that the conventional superimposed training scheme outperforms the conventional one both in terms of bit error rate and outage probability. We also prove that in data-dependent superimposed training, a suboptimum allocation of power between pilots and data could deteriorate the bit error rate performance, and propose an optimal power allocation scheme.

The remainder of this paper is as follows: In the next section, we introduce the system model. After that, we review the channel estimation process for the conventional and the data-dependent superimposed training. Then, we derive a lower bound of the bit error rate and a closed-form for the outage probability. After that, we determine the optimal allocation of power for the data-dependent superimposed training scheme. We prove that with this optimal allocation, the data-dependent superimposed training scheme is always more performant than the conventional scheme. Finally, simulation results are then provided to validate the analytical derivation.

Notation: Subscripts ^{H} and $^\#$ denote hermitian and pseudo-inverse operators. The statistical expectation and the Kronecker product are denoted by $\mathcal{E}$ and $\otimes$. The $(K \times K)$ identity matrix is denoted by $\mathbf{I}_K$, and the $(Q \times Q)$ matrix of all ones by $\mathbf{1}_Q$. The (i,j)$^{\text{th}}$ entry of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}_{i,j}$.

3. SYSTEM MODEL

In this paper, we consider a MIMO wireless system operating over K transmit antennas and M receive antennas with $M \geq K$. The channel is modeled as a spatially uncorrelated frequency flat fading channel that is time invariant over a single block. We denote by N the block length. At the receiver, the received signal for each block could be expressed as:

$$\mathbf{Y} = \mathbf{HS} + \mathbf{V},$$

where the $(K \times N)$ and $(M \times N)$ matrices $\mathbf{S}$ and $\mathbf{Y}$ are the transmitted and received matrix blocks. $\mathbf{H}$ is the $(M \times K)$ channel matrix to be estimated with independent and identically distributed (iid) gaussian variables with zero mean and variance $\frac{1}{K}$, and $\mathbf{V}$ is $(M \times N)$ additive noise matrix whose entries are iid zero mean with variance ${\sigma_v}^2$.

4. CHANNEL ESTIMATION

4.1. Conventional superimposed training

In the conventional superimposed training (ST) scheme, a known training sequence $\mathbf{P}$ is added to the data matrix $\mathbf{W}$, that is $\mathbf{S} = \mathbf{W} + \mathbf{P}$

The received signal is given by:

$$\mathbf{Y} = \mathbf{H}(\mathbf{P} + \mathbf{W}) + \mathbf{V},$$

where the entries of $\mathbf{W}$ are assumed to be iid with variance σ_w^2. Therefore, the total power σ_T^2 verifies:

$$\sigma_T^2 = \sigma_w^2 + \sigma_P^2,$$

where σ_P^2 denotes the training power.

The receiver estimates the channel by treating $\mathbf{HW}$ as an additive noise term. Hence the least squares channel matrix estimate is given by:

$$\begin{aligned}\widehat{\mathbf{H}} &= \mathbf{YP}^{\mathrm{H}}\left(\mathbf{PP}^{\mathrm{H}}\right)^{-1} \\ &= \mathbf{H}+\mathbf{HWP}^{\mathrm{H}}\left(\mathbf{PP}^{\mathrm{H}}\right)^{-\mathbf{1}}+\mathbf{VP}^{\mathrm{H}}\left(\mathbf{PP}^{\mathrm{H}}\right)^{-1}.\end{aligned}$$

Let $\mathbf{\Delta H}=\mathbf{H}-\widehat{\mathbf{H}}$ denote the channel estimation error matrix. Thus the mean square error (MSE) is given by:

$$\begin{aligned}\text{MSE} &= \operatorname{tr}\left(\mathcal{E}\left[\mathbf{\Delta H \Delta H}^{\mathrm{H}}\right]\right) \\ &= \operatorname{tr}\left(\left(\mathbf{PP}^{\mathrm{H}}\right)^{-1}\right)\left(\sigma_w^2 \operatorname{tr}(\mathbf{HH}^{\mathrm{H}})+M\sigma_v^2\right). \quad (1)\end{aligned}$$

It is shown in [1], [2] that the training matrix which minimizes the MSE subject to a fixed power σ_P^2 must verify:

$$\mathbf{PP}^{\mathrm{H}}=N\sigma_P^2\mathbf{I}_K.$$

Thus, the expression for the MSE becomes:

$$\text{MSE}=\frac{K}{N\sigma_P^2}\left(\sigma_w^2\operatorname{tr}(\mathbf{HH}^{\mathrm{H}})+M\sigma_v^2\right).$$

Note that the estimation errors always exist even if the additive noise is not present. This is particularly due to the presence of the unknown data that acts like an extra source of noise during the channel estimation step.

4.2. Data-dependent superimposed training

In [2], M. Ghogho and A. Swami propose to introduce a linear distorsion to the data matrix at the transmitter so as to ensure that the estimation error is independent from the unkown data. They suggest to add a perturbation matrix to the data matrix that is given by:

$$\mathbf{E}=-\mathbf{WJ},$$

where $\mathbf{J}=\frac{1}{Q}\mathbf{1}_Q\otimes\mathbf{I}_K$ and $Q=\frac{N}{K}$.

The received signal at each block is therefore given by:

$$\mathbf{Y}=\mathbf{HW}(\mathbf{I}_N-\mathbf{J})+\mathbf{HP}+\mathbf{V}$$

Hence, the total power is split between pilots and data as follows:

$$\sigma_T^2=\underbrace{\sigma_w^2}_{\text{data power}}-\underbrace{\frac{\sigma_w^2}{Q}}_{\text{distorsion power}}+\underbrace{\sigma_{P'}^2.}_{\text{training power}}$$

The design of the training matrix should meet the following conditions:

$$\begin{cases}\mathbf{PP}^{\mathrm{H}}=N\sigma_{P'}^2\mathbf{I}_K\\ \mathbf{P}^{\mathrm{H}}\mathbf{P}=N\sigma_{P'}^2\mathbf{J}.\end{cases}$$

It is easy to verify that under these conditions the training matrix $\mathbf{P}$ is orthogonal to the matrix $\mathbf{I}_N-\mathbf{J}$, (that is $(\mathbf{I}_N-\mathbf{J})\mathbf{P}^{\mathrm{H}}=\mathbf{0}$). Therefore, when mutiplying the received signal on the right side by $\mathbf{P}^{\mathrm{H}}$, the obtained result is independent from the unkown data.

The channel least square estimate is then given by:

$$\begin{aligned}\widehat{\mathbf{H}} &= \mathbf{YP}^{\mathrm{H}}\left(\mathbf{PP}^{\mathrm{H}}\right)^{-1} \\ &= \mathbf{H}+\mathbf{VP}^{\mathrm{H}}\left(\mathbf{PP}^{\mathrm{H}}\right)^{-1}.\end{aligned}$$

Thus the mean square error has the following expression:

$$\begin{aligned}\text{MSE} &= M\sigma_v^2\operatorname{tr}\left(\left(\mathbf{PP}^{\mathrm{H}}\right)^{-1}\right) \\ &= \frac{KM\sigma_v^2}{N\sigma_{P'}^2}. \quad (2)\end{aligned}$$

5. DATA DETECTION

5.1. Conventional superimposed training

For the conventional superimposed training, the zero forcing estimate of the transmitted data matrix writes as:

$$\begin{aligned}\widehat{\mathbf{W}} &= \left(\widehat{\mathbf{H}}\right)^{\#}\left(\mathbf{Y}-\widehat{\mathbf{H}}\mathbf{P}\right) \\ &= \left(\widehat{\mathbf{H}}\right)^{\#}\mathbf{H}\left(\mathbf{W}+\mathbf{P}\right)+\left(\widehat{\mathbf{H}}\right)^{\#}\mathbf{V}-\mathbf{P}. \quad (3)\end{aligned}$$

Assuming that the channel estimation error is small, the pseudo-inverse of the estimated matrix can be approximated by the linear part of the Taylor expansion as:

$$\left(\widehat{\mathbf{H}}\right)^{\#}=\mathbf{H}^{\#}-\mathbf{H}^{\#}\left(\mathbf{\Delta H}\right)\mathbf{H}^{\#} \quad (4)$$

where $\Delta\mathbf{H}$ denotes the channel estimation error matrix.

Substituting (4) into (3), the zero forcing estimate of the transmitted matrix can be further expressed as:

$$\widehat{\mathbf{W}}=\left(\mathbf{I}_K-\mathbf{H}^{\#}\Delta\mathbf{H}\right)(\mathbf{W}+\mathbf{P})+\left(\mathbf{I}_K-\mathbf{H}^{\#}\Delta\mathbf{H}\right)\mathbf{H}^{\#}\mathbf{V}-\mathbf{P}.$$

Hence, the effective post-processing noise $\Delta\mathbf{W}=\widehat{\mathbf{W}}-\mathbf{W}$ could be written as:

$$\begin{aligned}\Delta\mathbf{W} &= -\mathbf{W}+\left(\mathbf{I}_K-\mathbf{H}^{\#}\Delta\mathbf{H}\right)(\mathbf{W}+\mathbf{P}) \\ &\quad +\left(\mathbf{I}_K-\mathbf{H}^{\#}\Delta\mathbf{H}\right)\mathbf{H}^{\#}\mathbf{V}-\mathbf{P}.\end{aligned}$$

Assuming that the channel estimation error is uncorrelated with noise and data, we find that:[1]

$$\begin{aligned}\mathcal{E}\left(\Delta\mathbf{W}\mathbf{\Delta W}^{\mathrm{H}}\right)=&\left(\tfrac{\sigma_w^4 K}{\sigma_P^2}+K\sigma_w^2+\tfrac{\sigma_v^2\sigma_w^2}{\sigma_P^2}\operatorname{tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right)\mathbf{I}_K \\ +&\left(\tfrac{K\sigma_w^2\sigma_v^2}{\sigma_P^2}+K\sigma_v^2+N\sigma_v^2+\tfrac{\sigma_v^4}{\sigma_P^2}\operatorname{tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right)\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}. \quad (5)\end{aligned}$$

Assuming that the noise level is low, and that the probability that $\operatorname{tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}$ being large is very small [3], (5) becomes:

$$\begin{aligned}\mathcal{E}\left(\Delta\mathbf{W}\Delta\mathbf{W}^{\mathrm{H}}\right) \cong &\left(\frac{\sigma_w^4 K}{\sigma_P^2}+K\sigma_w^2\right)\mathbf{I}_K \\ &+\left(\frac{K\sigma_w^2\sigma_v^2}{\sigma_P^2}+K\sigma_v^2+N\sigma_v^2\right)\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}. \quad (6)\end{aligned}$$

[1] We omit all the proofs in this paper due to space limitation

5.2. Data-dependent superimposed training

Performing the same approximations as in the previous section, we find similarly that the variance of the effective post-processing noise in the data-dependent pilot scheme is:

$$\begin{aligned}\mathcal{E}\left(\Delta\mathbf{W}\boldsymbol{\Delta}\mathbf{W}^{\mathrm{H}}\right) = & \tfrac{N\sigma_w^2}{Q}\mathbf{I}_K \\ & + \left((N+K)\sigma_v^2 + \tfrac{K(Q-1)\sigma_w^2\sigma_v^2}{Q\sigma_{P'}^2}\right)\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}.\end{aligned}$$

6. PERFORMANCE ANALYSIS

6.1. A lower bound for the bit error rate

6.1.1. Conventional superimposed training

According to (6), the post-processing SNR on the kth stream can be expressed as:

$$\gamma_k = \frac{1}{\alpha_c + \beta_c\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{kk}}$$

where

- $\alpha_c = \frac{K}{N} + \frac{K\sigma_w^2}{N\sigma_P^2}$,
- $\beta_c = \frac{(K+N)\sigma_v^2}{N\sigma_w^2} + \frac{K\sigma_v^2}{N\sigma_P^2}$.

As the diagonal elements of $\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}$ are always positive, the SNR at the k^{th} branch could be maximized by:

$$\gamma_k \leq \frac{1}{\beta_c\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{kk}} = \gamma_c.$$

From [4] and [5], we know that $\frac{1}{K}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{kk}$ is a chisquare distributed random variable with $2(M-K+1)$ degrees of freedom. Thus, the probability density function of $\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{kk}$ can be expressed as:

$$f(x) = \frac{K^{M-K+1}x^{M-K}e^{-Kx}}{(M-K)!},$$

and consequently, the probability density function of γ_c is:

$$f_\gamma(x) = \frac{K^{M-K+1}\beta_c^{M-K+1}}{(M-K)!}x^{M-K}e^{-Kx\beta_c}.$$

This expression leads to the BER value:

$$\begin{aligned}\mathrm{BER}_c &= \int_0^\infty Q(\sqrt{2x})f_\gamma(x)dx \\ &\cong \frac{1}{2}\left[1-\mu_c\sum_{k=0}^{M-K}C_{2k}^k\left(\frac{(1-\mu_c^2)}{4}\right)^k\right] \qquad (7)\end{aligned}$$

where $\mu_c = \sqrt{\frac{1}{K+\beta_c}}$.

6.1.2. Data-dependent superimposed training

Following the same reasoning as in the previous section, we find that the post-processing SNR on the kth stream can be expressed as:

$$\gamma_k = \frac{1}{\alpha_d + \beta_d\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{kk}} \leq \frac{1}{\beta_d\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{kk}}$$

where $\alpha_d = \frac{1}{Q} = \frac{K}{N}$ and $\beta_d = \frac{\sigma_v^2}{\sigma_w^2}(1+\frac{K}{N}) + \frac{(N-K)\sigma_v^2}{N^2\sigma_{P'}^2}$.

In the same way, the lower bound of the BER is:

$$\mathrm{BER}_d \cong \frac{1}{2}\left[1-\mu_d\sum_{k=0}^{M-K}C_{2k}^k\left(\frac{(1-\mu_d^2)}{4}\right)^k\right]$$

where $\mu_d = \sqrt{\frac{1}{K+\beta_d}}$.

6.2. Optimization over the power allocation for the data-dependent superimposed training scheme

The power allocation between pilots and data has a great impact on the bit error rate performance. On the one hand, to achieve high quality channel estimate, a high portion of power needs to be spent for training transmission, which leaves little power to data. On the other hand, if too little power is given to training, the channel estimation will be poor which also affects the bit error rate performance. In this section, we give the values of data and training power portions that minimize the bit error rate for the data-dependent superimposed training scheme. As α_d do not depend on the data power σ_w^2 or the training power $\sigma_{P'}^2$, minimizing β_d as a function of σ_w^2 and $\sigma_{P'}^2$ under the constraint of constant total power leads to maximizing the SNR at each channel realization, and thus the bit error rate. The optimal values of σ_w^2 and $\sigma_{P'}^2$ are given by:

$$\left\{\begin{array}{lcl}\sigma_w^2 &=& \frac{\sigma_T^2(N^2+KN)-\sigma_T^2N\sqrt{N+K}}{N^2-N-K^2+K} \\ \sigma_{P'}^2 &=& \frac{\sigma_T^2\left(N\sqrt{N+K}-KN-N+K^2+K\right)}{N^2-N-K^2+K}.\end{array}\right.$$

In the conventional superimposed training scheme, the values of σ_w^2 and σ_P^2 that minimize β_c may not be optimal in the sense that they do no necessarily minimize the bit error rate.

6.3. BER performance of the optimum data-dependent superimposed training scheme compared to the conventional one

In this section, we prove that the data-dependent superimposed training scheme with optimal allocation of data and training power always outperforms the conventional superimposed training scheme. In fact, we show that for any possible values of data and training power, the SNR of the conventional superimposed training scheme is always lower than that of the optimum data-dependent superimposed training scheme. Actually, we prove the following:

$\forall\sigma_w^2 \in [0, \sigma_T^2]$ such that $\sigma_w^2 + \sigma_P^2 = \sigma_T^2$ and
$\sigma_w^2(1-\frac{1}{Q}) + \sigma_{P'}^2 = \sigma_T^2$, we have $\beta_c > \beta_d$ and $\alpha_c > \alpha_d$.

In fact, expressing β_c and β_d in terms of the total and data power leads to:

$$\beta_c(\sigma_w^2) = \frac{(K+N)\sigma_v^2}{\sigma_w^2} + \frac{K\sigma_v^2}{N(\sigma_T^2-\sigma_w^2)}$$
$$\beta_d(\sigma_w^2) = \frac{(K+N)\sigma_v^2}{\sigma_w^2} + \frac{(N-K)\sigma_v^2}{N^2(\sigma_T^2-\sigma_w^2(1-\frac{1}{Q}))}. \quad (8)$$

After simplification, $\beta_d - \beta_c$ writes as:

$$\beta_d - \beta_c = \frac{(N-K-KN)\sigma_v^2(\sigma_T^2-\sigma_w^2) - \frac{KN\sigma_w^2\sigma_v^2}{Q}}{N^2(\sigma_T^2-(1-\frac{1}{Q})\sigma_w^2)(\sigma_T^2-\sigma_w^2)}$$

which is an increasing function of σ_w^2.

As $\beta_d(0)-\beta_c(0) < 0$ and $\beta_d(\sigma_T^2)-\beta_c(\sigma_T^2) < 0$, we conclude that for any value of $\sigma_w^2 \in [0, \sigma_T^2]$, we have $\beta_c(\sigma_w^2) > \beta_d(\sigma_w^2)$.

Since $\alpha_c > \alpha_d$, we conclude that the data-dependent superimposed training scheme outperforms the conventional one when having the same data power. Therefore, the optimum data-dependent superimposed training scheme always outperforms the conventional one.

6.4. Closed-form for outage probability

Outage probability is defined as the probability that the post-processing SNR falls below some specified threshold:

$$\mathrm{P_{out}}(\gamma_{th}) = \mathrm{Prob}\,[0 \le \gamma \le \gamma_{th}]\,.$$

We could prove through simple calculations that the probability density function of the SNR is:

$$f_\gamma(x) = \frac{\beta_s^{M-K+1} x^{M-K} K^{M-K+1}}{(M-K)!(1-\alpha_s x)^{M-K+2}} e^{\frac{-K\beta_s x}{1-\alpha_s x}} \mathbf{1}_{\{0\le x\le \frac{1}{\alpha_s}\}}$$

where 's' is equal to 'c' to refer to a conventional superimposed training scheme or 'd' to refer to a data-dependent superimposed training scheme.

After straighforward calculations, we found that:

$$\mathrm{P_{out}}(\gamma_{th}) = \begin{cases} \frac{\Gamma(M-K+1, \frac{K\beta_s\gamma_{th}}{1-\alpha_s\gamma_{th}})}{(M-K)!} & \text{if } 0 \le \gamma_{th} \le \frac{1}{\alpha_s} \\ 1 \text{ if } \gamma_{th} \ge \frac{1}{\alpha_s} \end{cases}$$

where Γ is the lower incomplete gamma function given by:

$$\Gamma(a,x) = \int_0^x t^{a-1}e^{-t}\mathrm{d}t.$$

We note that the outage probability reaches the upper limit when γ_{th} tends to $\frac{1}{\alpha_s}$. As $\alpha_c > \alpha_d$, we could deduce that the outage probability of a conventional superimposed training scheme reaches its limit before that of the superimposed data-dependent training scheme.

7. SIMULATION RESULTS

In this section, we assess through simulations that the data-dependent superimposed training scheme outperforms the conventional superimposed training one in terms of bit error rate and outage probability. We prove also that setting arbitrarily the amount of power allocated to data and training can affect drastically the performance of the data-dependent superimposed training scheme.

In all our simulations, we set $M = 4$, $K = 2$, $N = 500$ and $\sigma_T^2 = 1$. The SNR is defined as SNR $\triangleq \frac{\sigma_T^2}{\sigma_v^2}$.

7.1. Bit error rate

7.1.1. Bit error rate performance

Fig. 1 investigates the performance in terms of bit error rate for the two schemes with BPSK modulation.

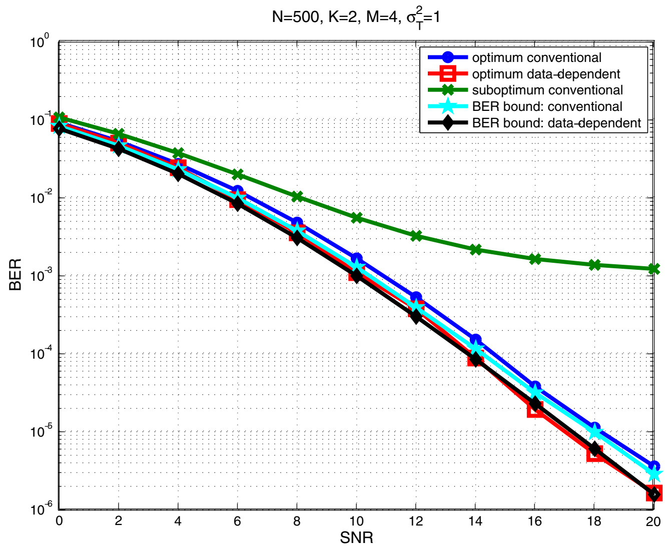

Fig. 1. BER performance of the conventional and the data-dependent superimposed training schemes.

In the legend, 'suboptimum conventional' corresponds to a conventional scheme having the same power as the optimum data dependent scheme.

In order to set the optimal power parameters for the conventional scheme, we measure, for each SNR value, the bit error rate for a range of power values between 0.05-0.99.

For each SNR value, we keep the amount of data power that optimizes the bit error rate performance.

We note that, when having optimal allocation of power, the conventional and the data-dependent scheme have almost the same performance.

7.1.2. Bit error rate degradation due to a suboptimum allocation of power

Fig. 2 compares in terms of bit error rate performance the optimum data-dependent superimposed training scheme with its suboptimum counterparts.

Comparing the suboptimum scheme having $\sigma_w^2 = 0.3$ and the optimum one, we could conclude that the optimum allocation of power could allow a considerable gain in SNR.

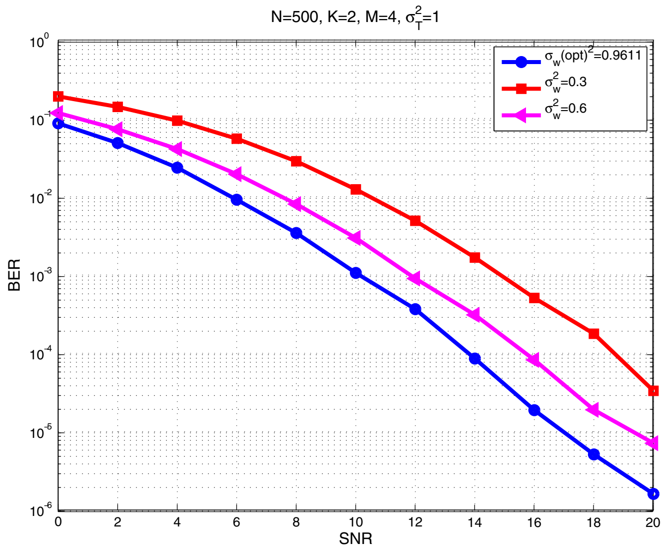

Fig. 2. Bit error rate degradation due to a suboptimum power allocation.

7.2. Outage probability

The performance in terms of outage probability is investigated in Fig. 3.

We note that the outage probability of the conventional superimposed training scheme reaches its upper limit before that of the data-dependent superimposed training scheme.

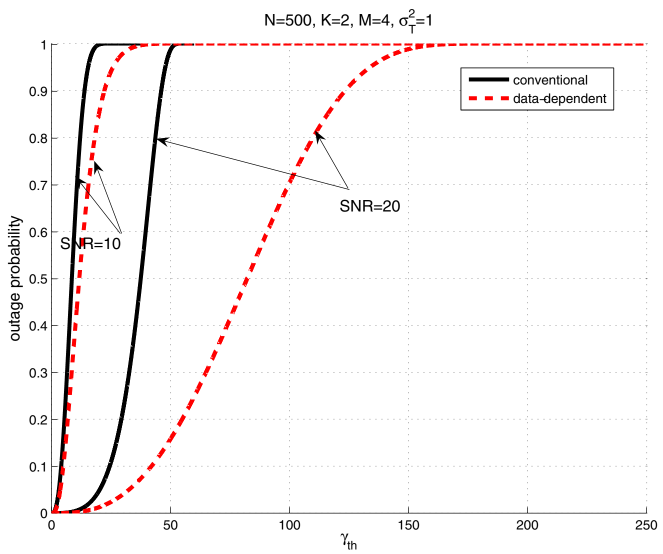

Fig. 3. Outage probability for the conventional and data-dependent superimposed training schemes.

8. CONCLUSION

In this paper, we analytically investigate the performance of the conventional and the data-dependent superimposed training schemes over uncorrelated Rayleigh flat fading channels. We derive for the two schemes a lower bound of the bit error rate and a closed-form for the outage probability. We prove that the data-dependent superimposed training scheme always outerperforms the conventional one. We also determine the optimal allocation of power between data and training for the data-dependent superimposed scheme.

9. REFERENCES

[1] B.Hassibi and B.M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Trans. Information Theory*, 2003.

[2] M. Ghogho and A. Swami, "Channel Estimation for MIMO Systems Using Data-Dependent Superimposed Training," *Allerton Conference*, 2004.

[3] C. Wang, E. K. S. Au, R. Murch, and V. Lau, "Closed-Form Outage Proability and BER of MIMO Zero-Forcing Receiver in the Presence of Imperfect CSI," *IEEE International Workshop on Signal Processing Advances in Wireless Communications*, 2006.

[4] J. Winters, J.Salz, and R. Gitlin, "The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems," *IEEE Trans. Commun.*, 1994.

[5] D. A. Gore, R. W. Heath, and A. J. Paulraj, "Transmit Selection in Spatial Multiplexing Systems," *IEEE Communications Letters*, 2002.